4

DIGITAL & SOLID STATE APPENDICES

Digital multimeters

Disadvantages of analogue meters

In *Digital & Solid State Appendices 3* we noted that the accuracy of analogue multimeters is determined primarily by the moving-coil movement itself and the position of the pointer on the scale. Thus, if a moving-coil movement is, say, accurate to ±5% of FSD, and the pointer indicates a reading of 40% of its full scale indication (i.e. it indicates, for example, 4 V on a 10 V FSD scale) then the actual measured voltage may be within the range:

$$4 \pm (5\% \text{ of } 10) = 4 \pm 0.5\,\text{V}$$

which means the indicated voltage is accurate to ±12.5% of the actual voltage.

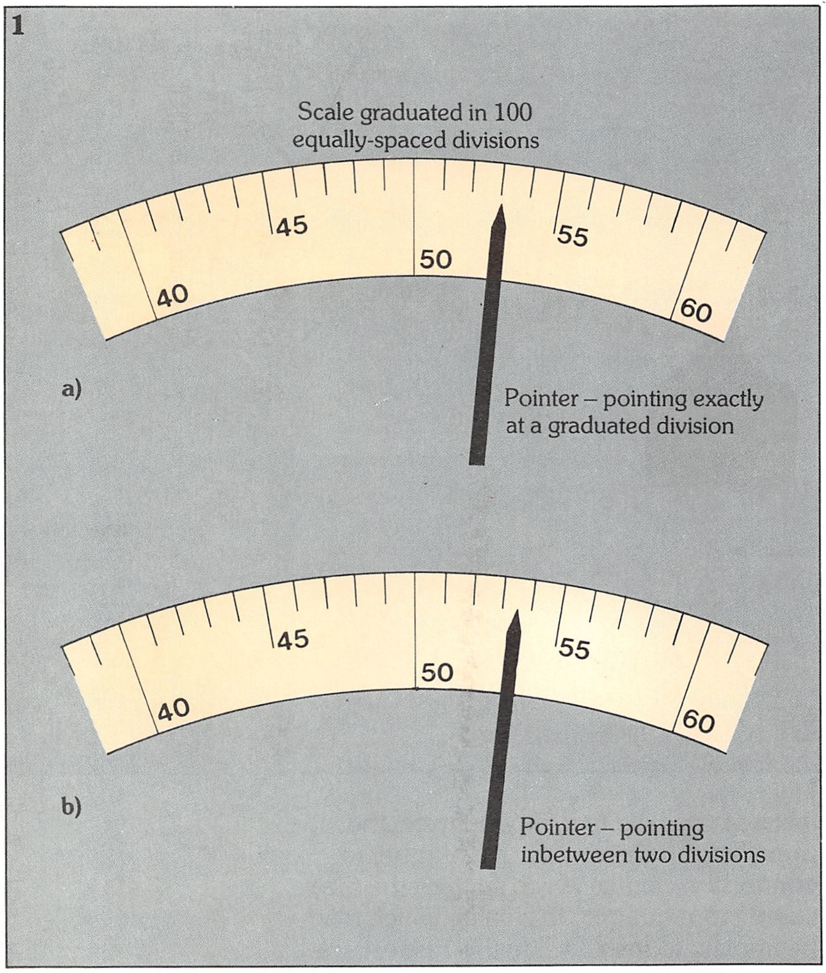
Another disadvantage of the use of analogue multimeters when measuring voltage arises if its input resistance is low, and the meter thus loads the circuit it is measuring, with the result that inaccurate measurements occur.

Yet another factor which may affect the accuracy of a measurement, taken with an analogue multimeter, is the **resolution** with which the indication is perceived by the observer. For example, the scale on an analogue multimeter may be graduated in divisions which are portions of the FSD value. If there are 100 divisions, and the pointer points *exactly* on one division (*figure 1a*) then the reading may be made at a resolution of one in 100: or 1%.

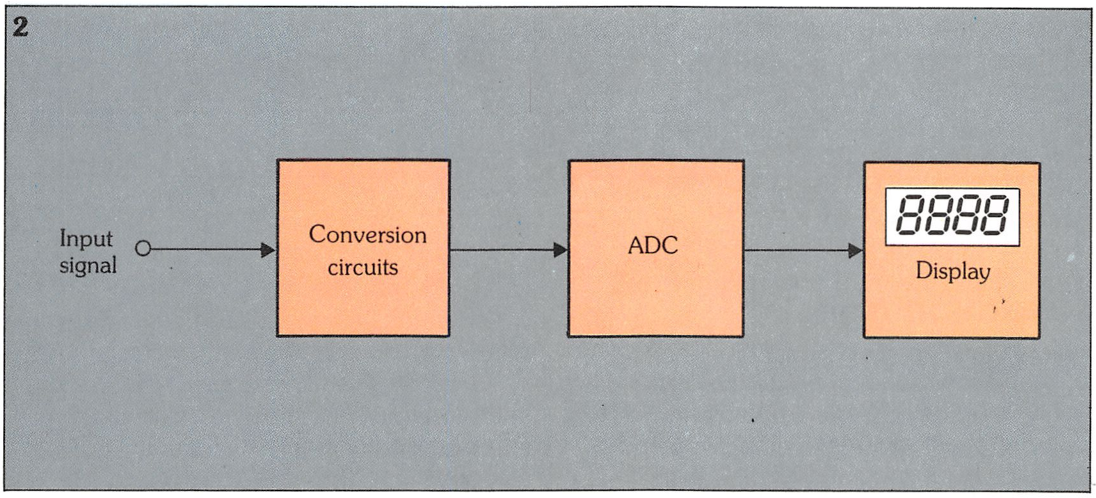
If the pointer points midway between two divisions (*figure 1b*) it might be possible to resolve the reading to one half of this, i.e. 0.5%. Chances are, however, that this resolution will not be possible because it is only a subjective judgement which enables us to define the pointer at *exactly* half-way between divisions.

A different type of multimeter, based on an entirely different principle to the moving-coil analogue multimeter, is the **digital multimeter**. In a digital multimeter, the voltage, current or resistance measurement is indicated in digital form as a numerical read-out. Inaccuracies due to meter movement are therefore eliminated and resolution of the reading is defined primarily by the resolution of the analogue-to-digital conversion which takes place within the instrument. Errors in indicated measurements are due mainly to component inaccuracies and poor calibration.

A possible block diagram of the main parts of a digital multimeter is shown in *figure 2* and we can see that three main stages are involved in the production of a

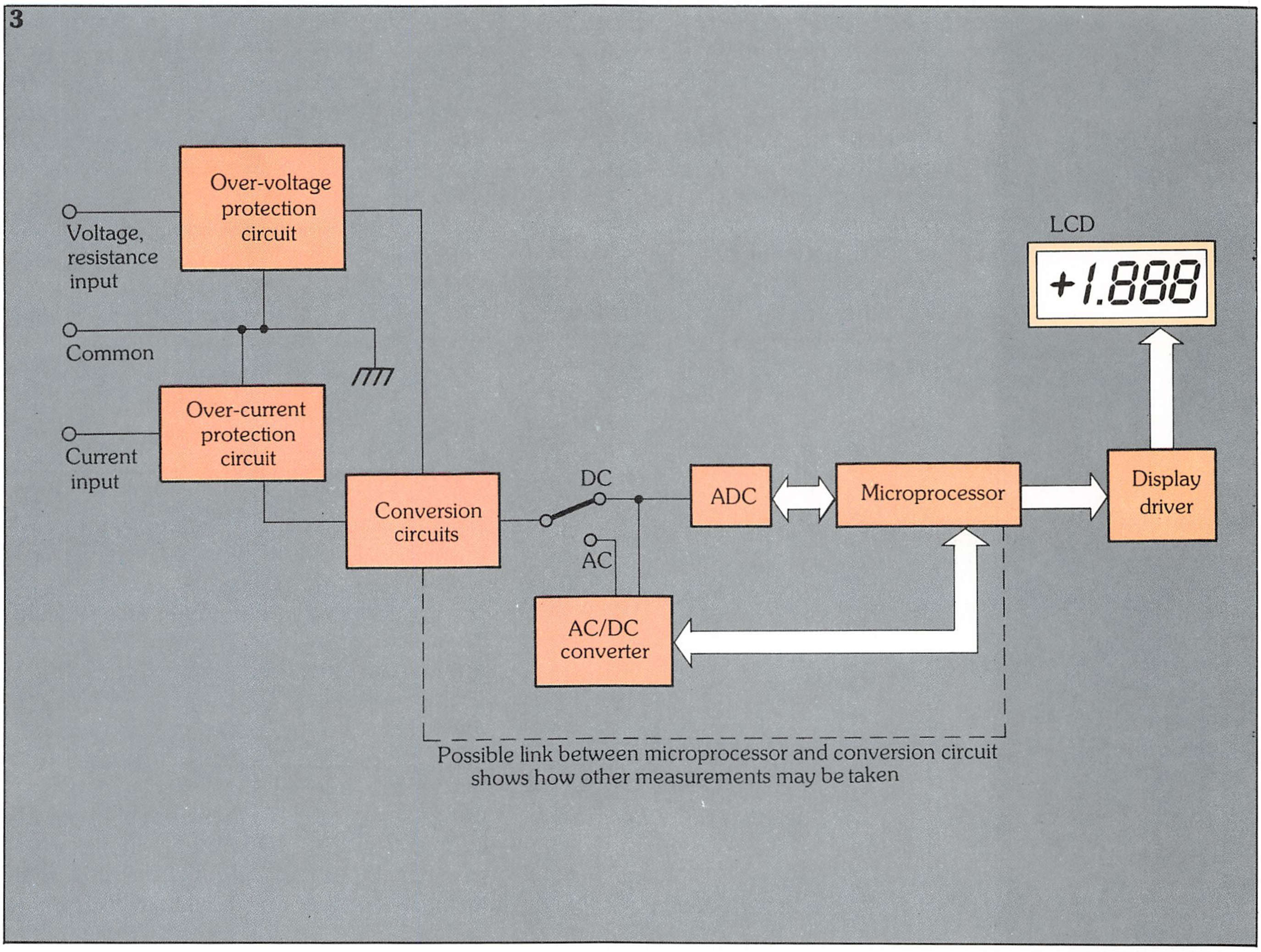


1. Analogue multimeter scale: (a) reading can be made at a resolution of 1%; **(b)** resolution of 0.5% is the best possible.



2. The three main stages in the production of a digital display of the input signal.

3. Detailed block diagram of a typical digital multimeter.



digital display of the input signal.

Central to the digital multimeter is the analogue-to-digital converter (ADC) which converts the sampled analogue signal to a digital form. Because an ADC is essentially a fixed range, voltage input device, the preceding stage consists of a number of conversion circuits which convert the measured input quantity to a value which may be applied to the ADC. In this respect, the

conversion circuits are similar in function to the shunts and multipliers of an analogue multimeter. Typically, a digital multimeter will indicate FSD when 200 mV, or 2 V is applied to the ADC, so the conversion circuits merely convert the applied voltage, current, resistance or units of whatever quantity we wish to measure to the required voltage.

The conversion circuits of a digital multimeter are always electronic and so the multimeter's input resistance is always extremely high, (when measuring voltage), preventing the possibility of excessively loading the measured circuit.

The majority of digital multimeters use seven-segment LCD displays – although LED displays are sometimes used. Low current consumption of the LCD types means that, with low current consumption ADC and conversion circuits (say, of CMOS origin), the complete digital multimeter has an overall current consumption of only a few μA. This is obviously an advantage if the multimeter is to be battery powered and hence portable.

The number of digits displayed depends on the type of digital multimeter, but generally between four and eight digits are common. A decimal point is usual, which moves along the display corresponding to the range displayed. Often, the most significant digit in the display is not of seven-segments but merely displays (or doesn't display) the number 1. If, say, the meter has three full seven-segment digits plus a most significant digit of this type, it is known as a **three-and-a-half digit display**. Maximum indication produced by a 3½ digit display is 1999.

A more detailed block diagram of a typical digital multimeter is shown in *figure 3* and we can see the three main stages of conversion circuits, ADC and display outlined. Also shown are circuits which protect the instrument in case of applications of large, potentially damaging voltages or currents. Protection circuits of this type are known as **over-range protection** circuits.

Generally, an applied over-range input signal produces a unique display (say, a 1 at the most significant digit and the remaining three digits blank) notifying the user of the over-range condition.

As digital multimeters are typically microprocessor controlled, it is a relatively simple task to use the microprocessor's computing ability to create measurement functions not usually found on analogue multimeters. Indications of temperature, capacitance, inductance, power etc., and mathematical calculations such as mean, average, standard deviation may be found, for example, on some digital multimeters.

Glossary

over-range protection	protection afforded in a digital multimeter, preventing damage if signals of large, potentially damaging, size are applied. Generally, if an over-range signal is applied, the digital multimeter display indicates the condition to the user
three-and-a-half digit	type of display commonly found in digital multimeters in which the most significant digit may only be indicated as a number 1 or as a blank. Maximum indication is thus 1999. (Displays with 4½, 5½, 6½ digits etc., are available)

Oscilloscopes

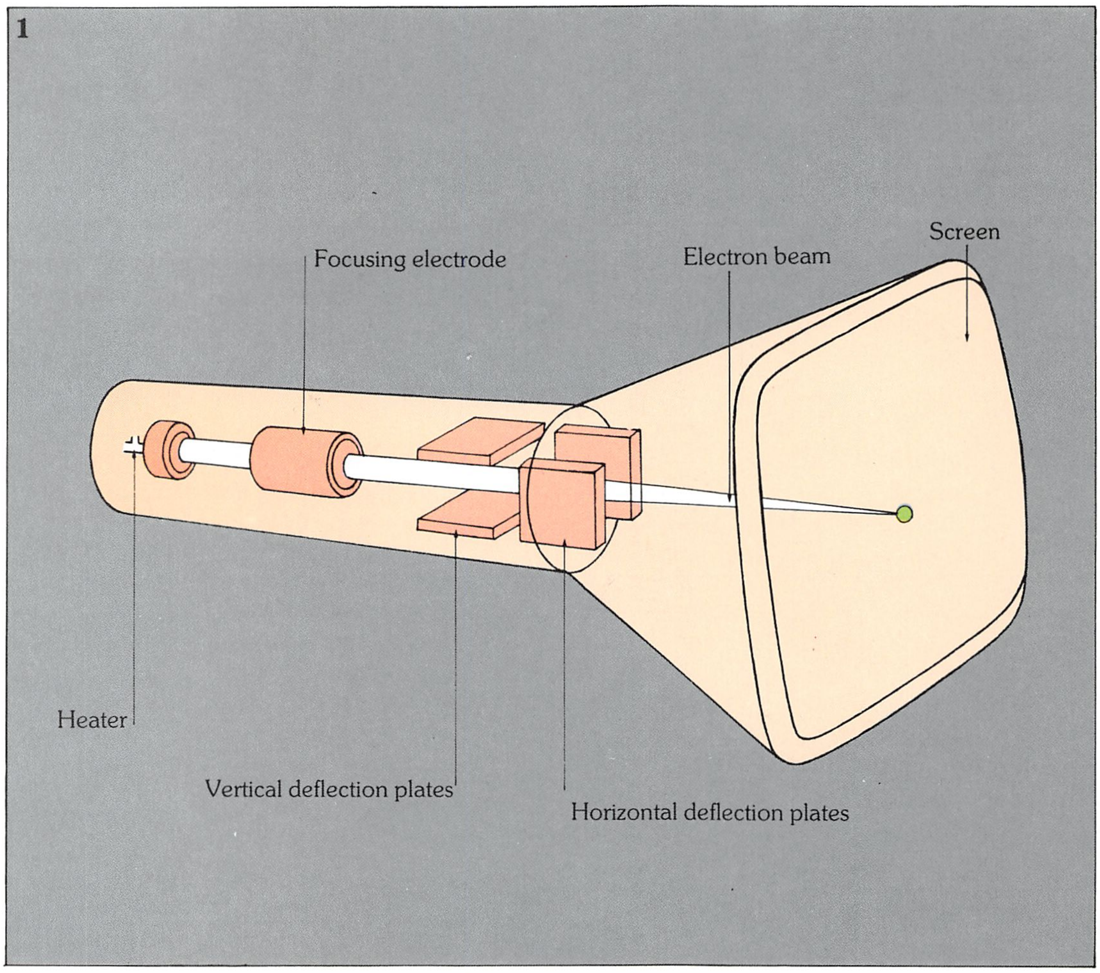
DIGITAL & SOLID STATE APPENDICES

An **oscilloscope** is a device which is capable of displaying a visual representation of a waveform. It may be used to display waveforms of voltage signals present in electronic circuits and is therefore useful as an item of test equipment in design, production and servicing of all kinds of other electronic equipment. For example, a television service engineer may use an oscilloscope to check the voltage signal waveforms present in a faulty television. By comparing those actually found with signal waveforms which *should* be present, the engineer may be able to isolate and repair the fault.

As the waveform displayed by the oscilloscope is a representation of the actual electrical signal producing the waveform, the oscilloscope may be used to *measure* the electrical signal. In this respect, the oscilloscope is used as a form of multimeter. However, it is far more versatile than this and we shall see why shortly.

The display

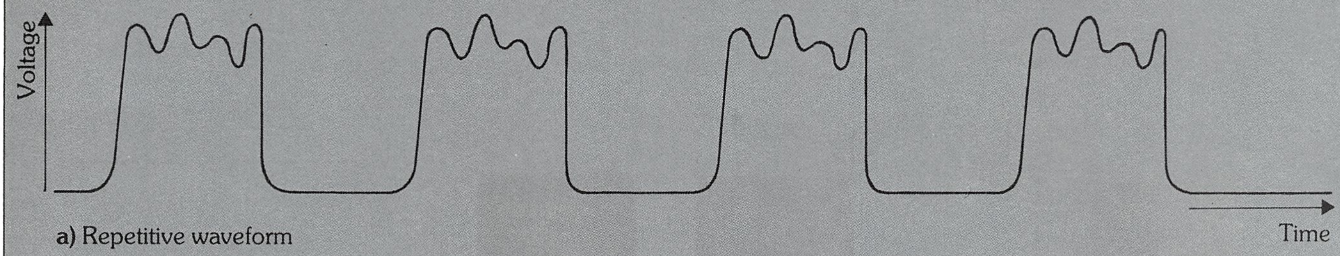
Generally, oscilloscope displays are of the cathode ray tube (CRT) electrostatic type, which we have already seen in *Solid State Electronics 23*. *Figure 1* shows the make-



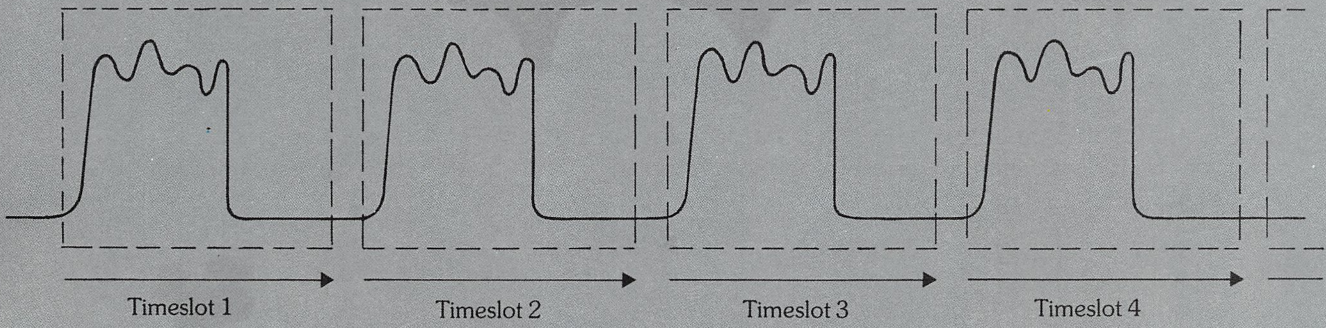
1. A typical oscilloscope CRT.

2. (a) Repetitive electrical signal applied to an oscilloscope; **(b)** waveform divided into equal timeslots; **(c)** waveform overlayed in successive timeslots.

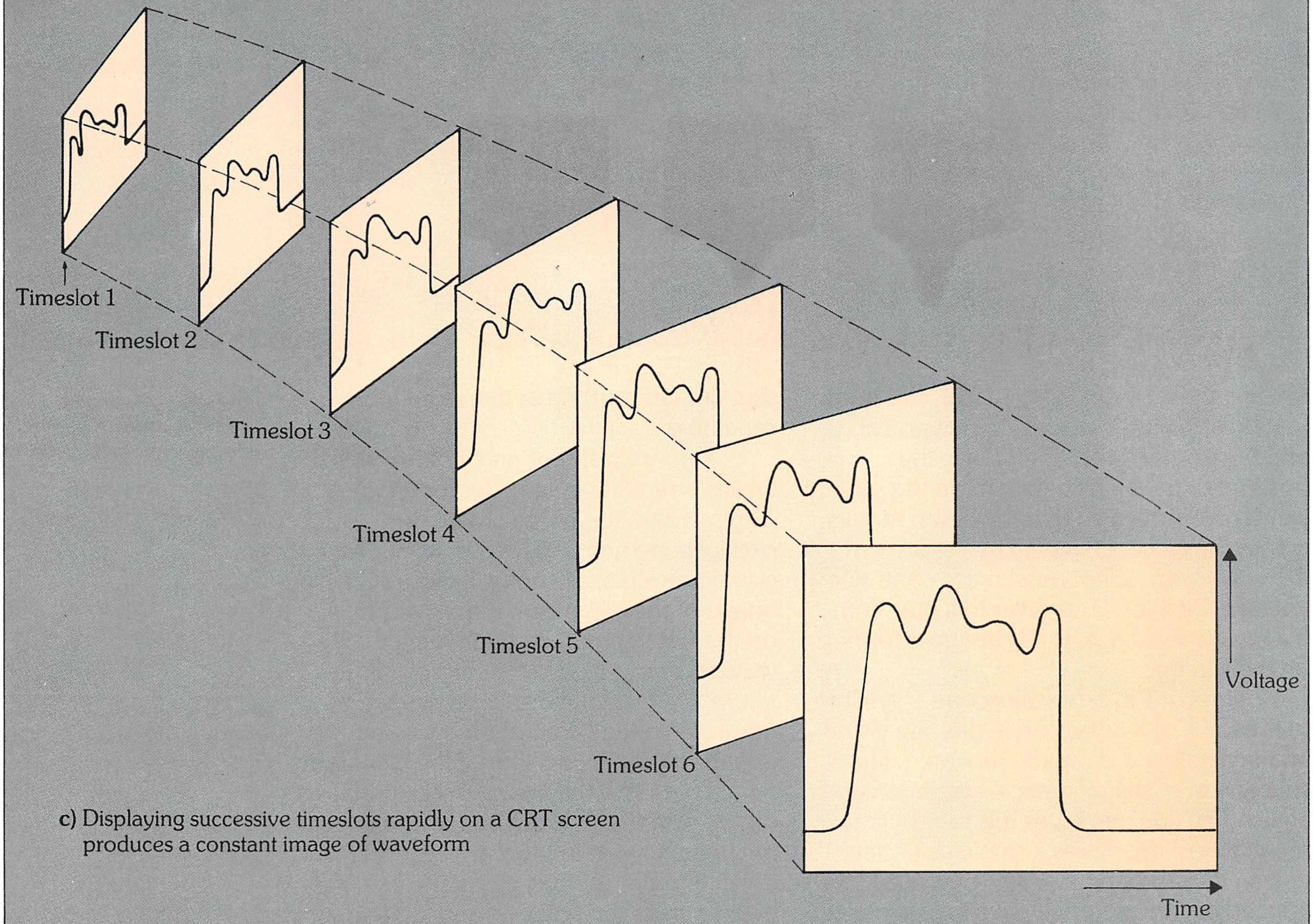
2



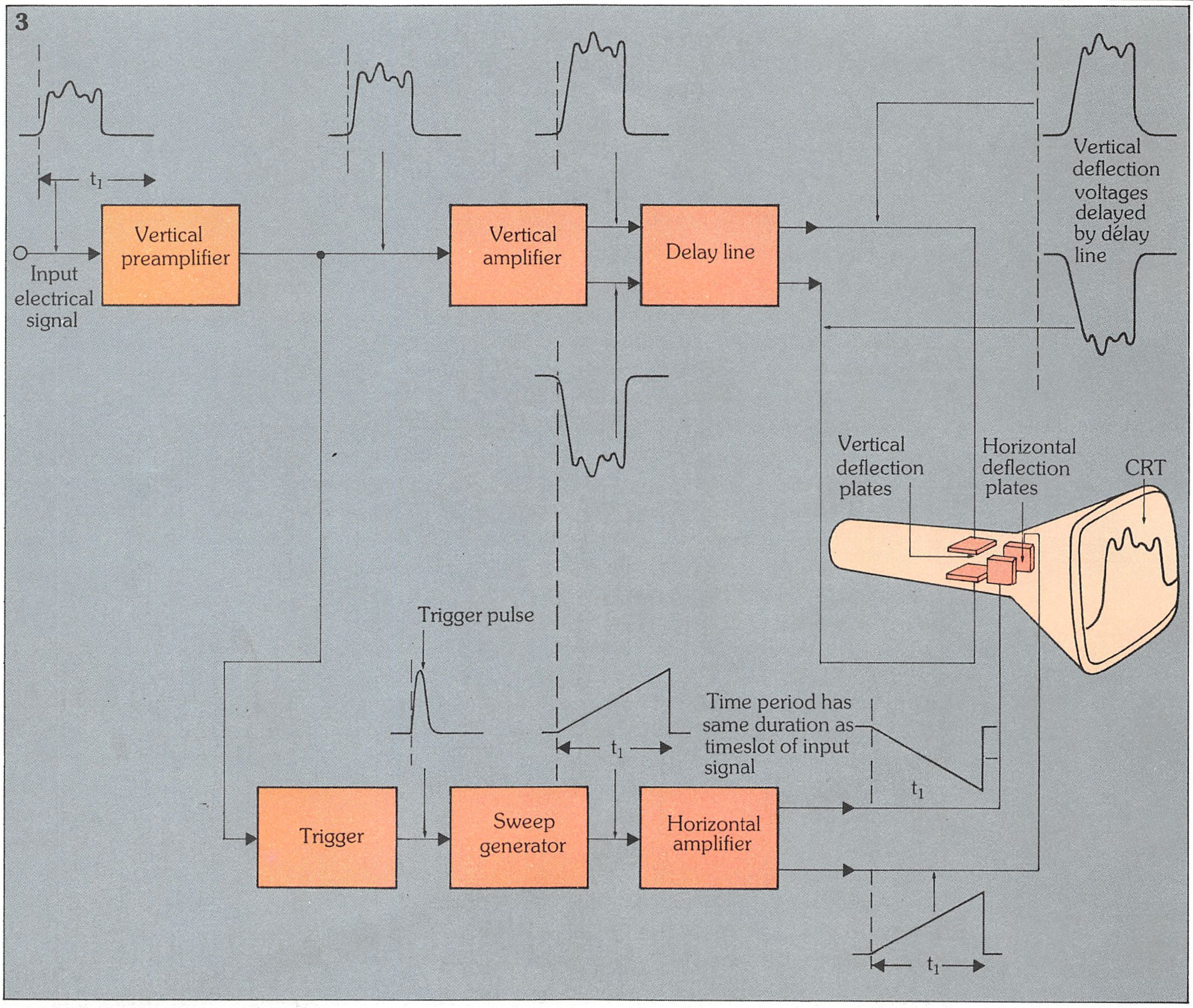
a) Repetitive waveform



b) Dividing into timeslots



c) Displaying successive timeslots rapidly on a CRT screen produces a constant image of waveform



3. **Action of a basic oscilloscope** which only displays one waveform.

up of a typical oscilloscope CRT. We know that by applying voltages to the vertical and horizontal deflection plates, the position of the electron beam on the screen can be controlled so that the beam moves across the screen in a way that represents the waveform. What must these controlling voltages be like? To answer this question we must consider *how* the image is formed on the screen.

The CRT in an oscilloscope is similar to the CRT of a television in that the visual image displayed is rapidly repeated, like a succession of still pictures, over and over again. The eye perceives this rapid repetition of images as a single stationary image, due to its persistence of vision. The applied electrical signal to a *basic* oscilloscope (more complex types are slightly different) must therefore be repetitive.

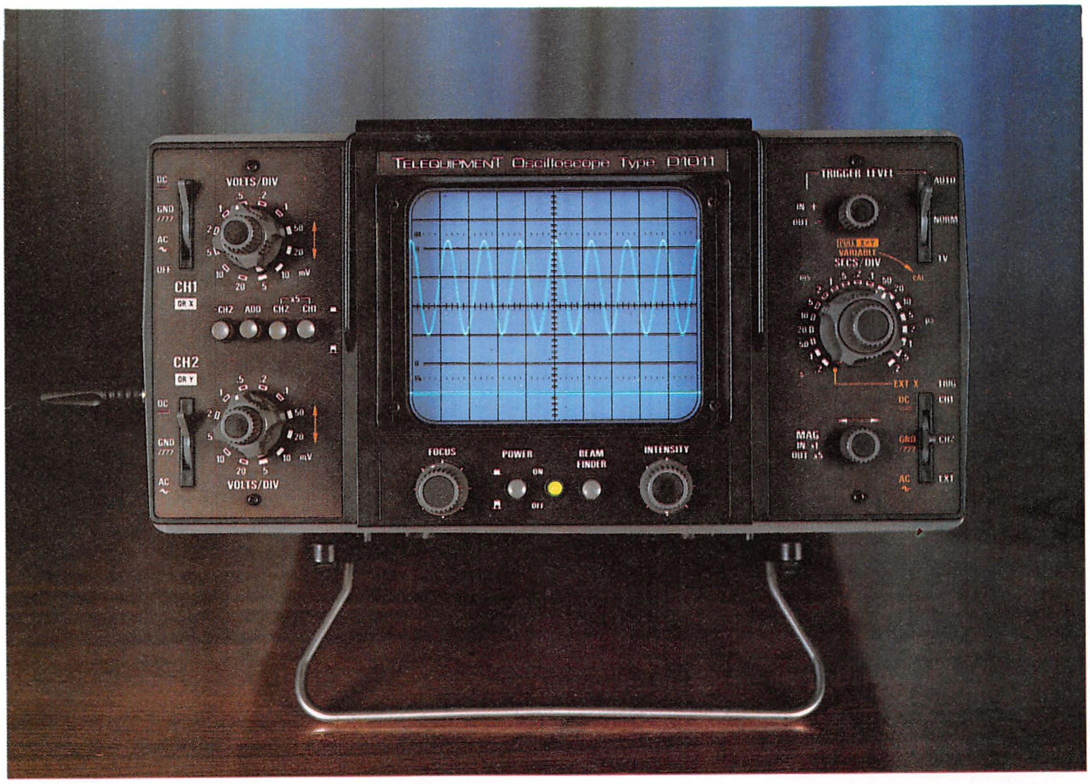
Figure 2a shows an electrical signal which is repetitive. In order for it to be displayed on the CRT screen, the oscilloscope first divides the waveform into equal timeslots, as shown in *figure 2b*. The image of the waveform is then displayed on the CRT screen by overlaying the waveform in successive timeslots, as in *figure 2c*. The final display is, in fact, a graph of voltage against time for an apparently single portion of the waveform.

To enable the image to be displayed on the screen therefore, we can see that the voltages applied to the horizontal deflection plates must move the electron beam across the screen, from left to right in what is known as a **sweep**, or a **timebase**;

the voltages applied to the vertical deflection plates, on the other hand, move the beam in a way which corresponds to the amplitude of the electrical signal measured. This is summarised in *figure 3* where the voltages to the horizontal deflection plates are seen as ramping voltages, whose time period, t_1, is the same time period as the timeslot of the input electrical signal measured. The voltages applied to the vertical deflection plates are amplified versions of the input electrical signal.

voltage to the required amplitude and the two resultant voltages (one the inverse of the other) are applied to the horizontal deflection plates.

Two stages of amplification are used to amplify the input electrical signal to the level required by the vertical deflection plates. The trigger circuit input is taken from the output of the first stage amplifier, known as the **vertical preamplifier**. The two output voltages of the second stage amplifier – the **vertical amplifier** – are passed through a delay line before they are



Right: a typical dual-trace oscilloscope.

Neville Miles

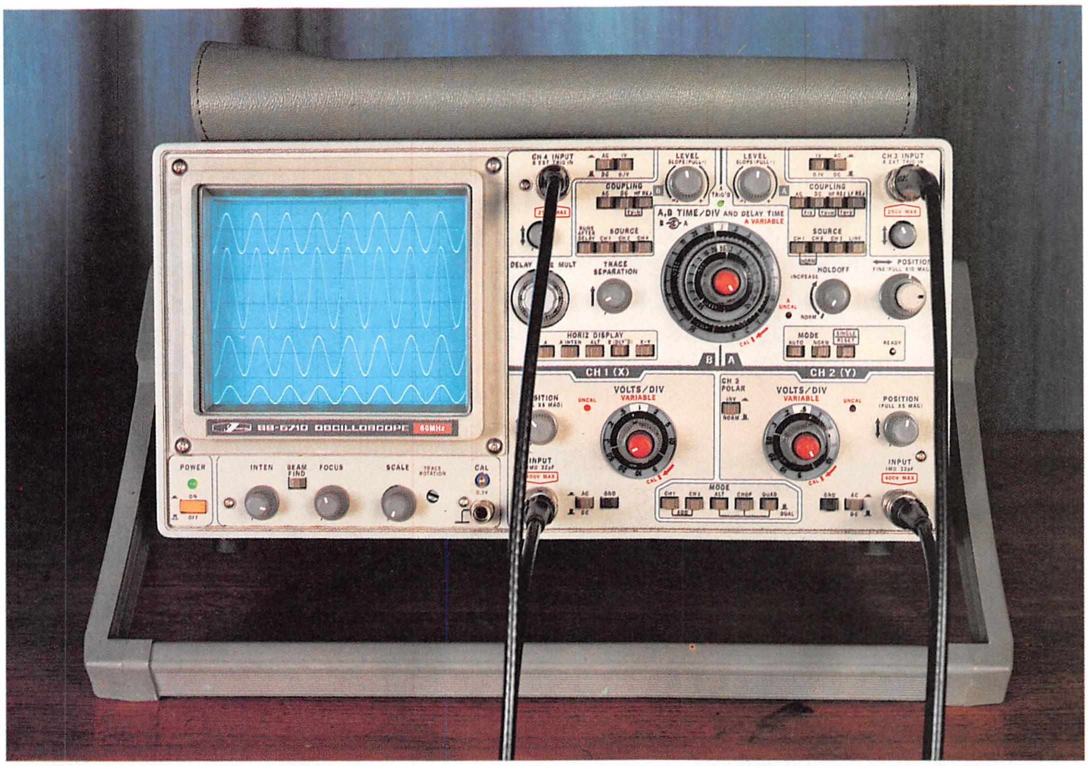
In order that the same part of the waveform is displayed by each sweep of the electron beam across the CRT screen, a **trigger** circuit is used to cause a pulse to occur as the input signal reaches a certain voltage. This circuit is adjustable so that the user can select at which point in the timeslot the sweep starts.

A **sweep generator**, also known as a **timebase generator**, is used to produce the required ramping voltage when the pulse is generated by the trigger circuit. The time period of the sweep generator is adjusted by the user to suit the required electrical signal timeslot. A **horizontal amplifier** amplifies the ramping sweep

applied to the vertical deflection plates. The delay line is required to compensate for delays inherent in the trigger circuit, sweep generator, and horizontal amplifier which would otherwise combine and prevent the leading edge of the waveform from being displayed.

The oscillscope shown in *figure 3* is a very basic type with only one trace, i.e. it can only display one waveform. Most practical oscillscopes have facilities such as **dual-trace display** where the electron beam is rapidly switched from one waveform to another so that the eye perceives a display of two waveforms.

Also, there are a number of controls



Neville Miles

Left: complex oscilloscope.

which allow the user to modify the displayed waveform. For example, the height of the displayed waveform may be changed by altering the gain of the vertical amplifier stages. This is usually effected by switching controlling resistors in and out of circuit. Sometimes, a variable resistor can also be used to control the gain between the fixed range steps. Controlling the gain of the vertical amplifier obviously allows signals of different amplitudes to be displayed by the oscilloscope.

The sweep time may also be user-controlled, in a similar way, enabling the oscilloscope to display waveforms corresponding to different signal frequencies.

Often, triggering mode may be selected so that the scope triggers in one of the following ways: on a positive-going edge; on a negative-going edge; automatically – without user intervention; from an external pulse; in synchrony with a television timebase (to allow display of television video signals); or from input signals applied to either trace of a dual-trace oscilloscope.

A photograph of a typical dual-trace oscilloscope, in which all of the controls producing these functions may be seen, is shown on page 1063. Other controls such as intensity and focus are also clearly visible. In contrast, a much more complex oscilloscope is shown above.

Using the oscilloscope

As mentioned earlier, the oscilloscope may be used to measure the applied electrical signal. Now that we have seen how the oscilloscope works, we can look at its use, in particular with reference to electronic measurements.

Compared with the multimeter, the main type of measuring instrument, the oscilloscope has the advantage of a high speed response – this is because the display (CRT) is much faster than the display of, say, a moving-coil meter. However, measurements taken using an oscilloscope have the disadvantage of their need to be calculated, from a known standard. For example, the sine wave on the oscilloscope of *figure 4* is exactly 4.5 divisions high (from peak-to-peak) and its wavelength is exactly 5 divisions. Now, let's say the vertical amplifier of the oscilloscope is set to 1 V per division and the horizontal amplifier is set to 1 ms per division. This means that as the peak-to-peak amplitude of the signal is 4.5 V, it's amplitude is thus 2.25 V, and its rms value

is thus 1.58 V rms.

Similarly, we may calculate the signal's frequency, by first calculating its period, T, as 5 ms. Now,

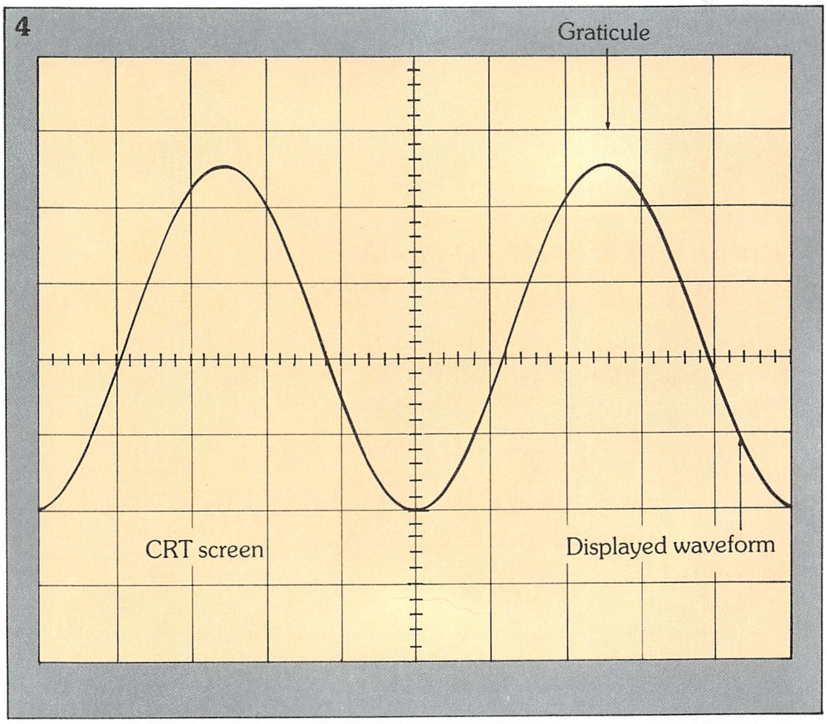
$$F = \frac{1}{T}$$

so the signal frequency is 200 Hz.

Therefore, by knowing the oscilloscope's vertical and horizontal amplification settings (read off from the front panel controls, which are calibrated) the oscilloscope may be used to measure the input signal amplitude and frequency.

Incidentally, the ten square by eight square grid, where each square is generally one centimetre square, is common to most oscilloscopes. The grid, called the **graticule**, is often scribed on a clear glass or plastic plate over the screen, and sometimes a system of side-lighting, through the covering glass or plastic plate, allows the graticule to effectively 'light up' if required by the user.

4. Measurements from an oscilloscope have to be calculated from a known standard.



Measurement of current

Being an electronic measurement device, the oscilloscope has a high-resistance input which means that it does not load the circuit under test. Essentially, therefore, it measures and displays voltages applied at the input.

If it is required to measure and display current, however, use may be made of a known value resistor, through which the current is passed. The oscilloscope may then be used to measure the voltage created across the resistor by the current. If, say, the resistor has a value of 1 Ω the voltage displayed and measured in volts on the oscilloscope may be equated directly as current in amps.

Special connecting leads which allow the oscilloscope to measure and display current are also available – these use a transformer to develop an output voltage across the secondary when an input current flows through the primary. Only AC currents can be measured, of course.

Storage oscilloscopes

The oscilloscopes we have considered have been based on the principle that the signal to be displayed is repetitive. By displaying one short segment of the signal, and repeating the display every time a similar segment occurs, the overall effect is of a stationary, non-changing signal.

Non-repetitive signals cannot, however, be displayed by this method. Take, for example, a microprocessor's interrupt request signal. We know that such a signal, upon receipt, may cause the microprocessor to process a specific routine before recommencing the program being processed before the interrupt. Such an interrupt request signal is probably not repetitive, and may well consist of a single pulse of only a few microseconds duration. To display this kind of non-repetitive signal a **storage oscilloscope** can be used – the display represents a *single* timeslot of the measured voltage, stored within the oscilloscope.

There are two main types of storage oscilloscope: one which depends on a special storage CRT; while the other uses digital devices to sample and store the applied voltage for later recall and display.

Of the two types, the **digital storage oscilloscope** (DSO) is the more versatile as the memorised voltage signal is in digital form, and its waveform may therefore be moved around on the screen, magnified, contracted, erased and then redisplayed etc., long after the recorded event has occurred. In contrast, the storage CRT

oscilloscope records a single event within the CRT itself, and the displayed waveform cannot be repositioned or altered in any way.

Summary

The oscilloscope is one of the most useful items of test equipment available, as it produces a displayed waveform of a measured voltage(s) in a voltage against time graph form. If the oscilloscope is adequately calibrated, measurements of amplitude and frequency of the measured voltage signals may be taken with reasonable accuracy.

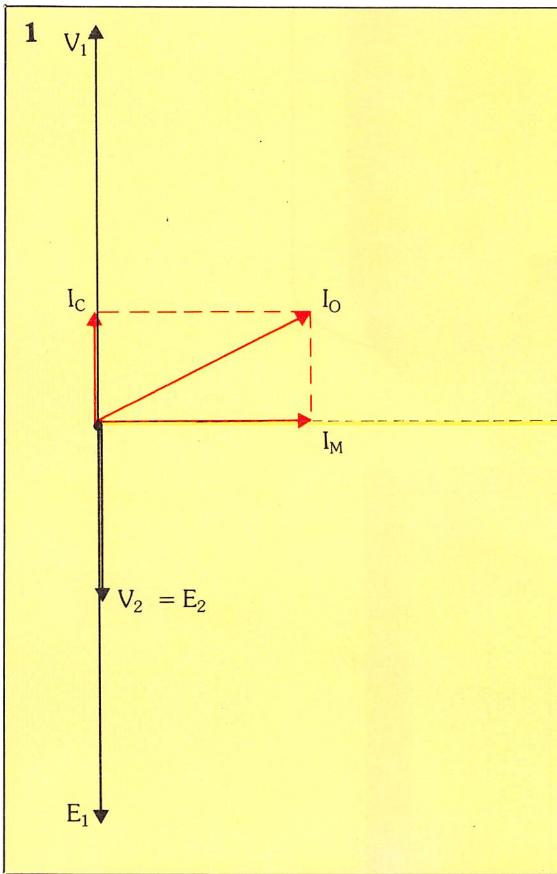
The basic types of oscilloscope can only display repetitive waveforms, but storage oscilloscopes are available which may display single events of very short duration.

Glossary

digital storage oscilloscope (DSO)	an oscilloscope which uses digital devices to sample and store applied voltage signals. The displayed waveform may therefore be of a single short event
dual-trace display	oscilloscope display which is capable of displaying two apparently separate waveforms. A dual-trace display is constructed using a single electron gun
graticule	the marked grid (usually 10 cm by 8 cm) on the face of an oscilloscope's CRT screen
oscilloscope	item of test equipment used to display waveform(s) representing applied voltage signal(s)
storage oscilloscope	an oscilloscope capable of storing one or more applied voltage signals. These signals may be represented by waveforms on the display until no longer required
sweep generator, timebase generator	circuit in an oscilloscope which generates a ramping voltage, to provide horizontal deflection of the electron beam, upon initiation by the trigger circuit
sweep, timebase	the horizontal deflection (i.e. sweep) of the electron beam, from left to right as viewed by the user
trace	a displayed waveform
trigger	circuit within an oscilloscope that helps to ensure each sweep of a displayed waveform commences at the same part of the measured voltage signal, thus ensuring a stable display of the repetitive signal

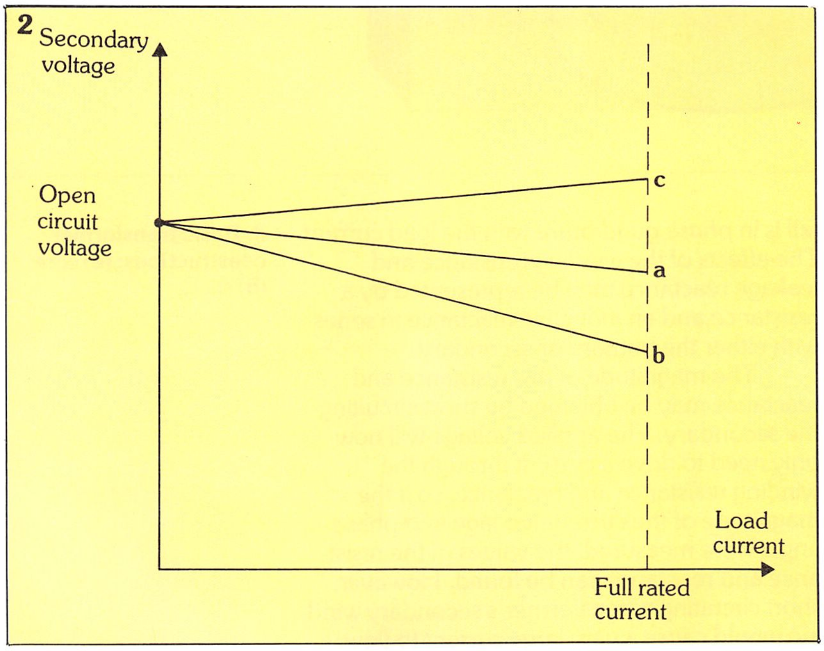
ELECTRICAL TECHNOLOGY

Transformers - regulation and construction



1. Phasor diagram showing how the rms values of V_1 and V_2 are related to peak flux.

2. Graph showing a transformer's variation of terminal voltage as current is increased for a resistive load.



In the last *Basic Theory Refresher*, we found that the rms value of the input and output voltages, V_1 and V_2, in a transformer are related to the peak flux in the following way:

$$V_1 = -E_1$$
$$= -4.44\, N_1 f \phi$$
$$V_2 = E_2$$
$$= 4.44\, N_2 f \phi$$

This can be shown by the phasor diagram in *figure 1*. Here the flux ϕ is shown in green and the input voltage leads it by 90°, while the output voltage lags it by 90°. If the transformer's secondary winding is open circuit, then no current will flow in it, and you might assume that the primary current is also zero. However, this is not quite correct. *Figure 1* shows us that a flux of maximum value ϕ is needed to maintain the terminal voltages. A certain magnetising force is necessary to keep the flux in the core, and this means that a magnetising current, I_M, must flow in the primary winding. This current will be in phase with the flux, and is shown in *figure 1*.

We also know that when the flux in an iron core is alternating, the core material is taken through a complete hysteresis loop, many times a second, resulting in a hysteresis loss. In addition to this, the alternating flux sets up eddy currents in the core's laminations, which cause an eddy current loss. Both of these losses combined, give the total core loss. This is a power loss, so we can represent it as a current, I_C. in phase with the primary voltage (*figure 1*). The two currents that flow when the transformer is unloaded may be combined by the parallelogram rule to give the **no-load** current, I_o, where:

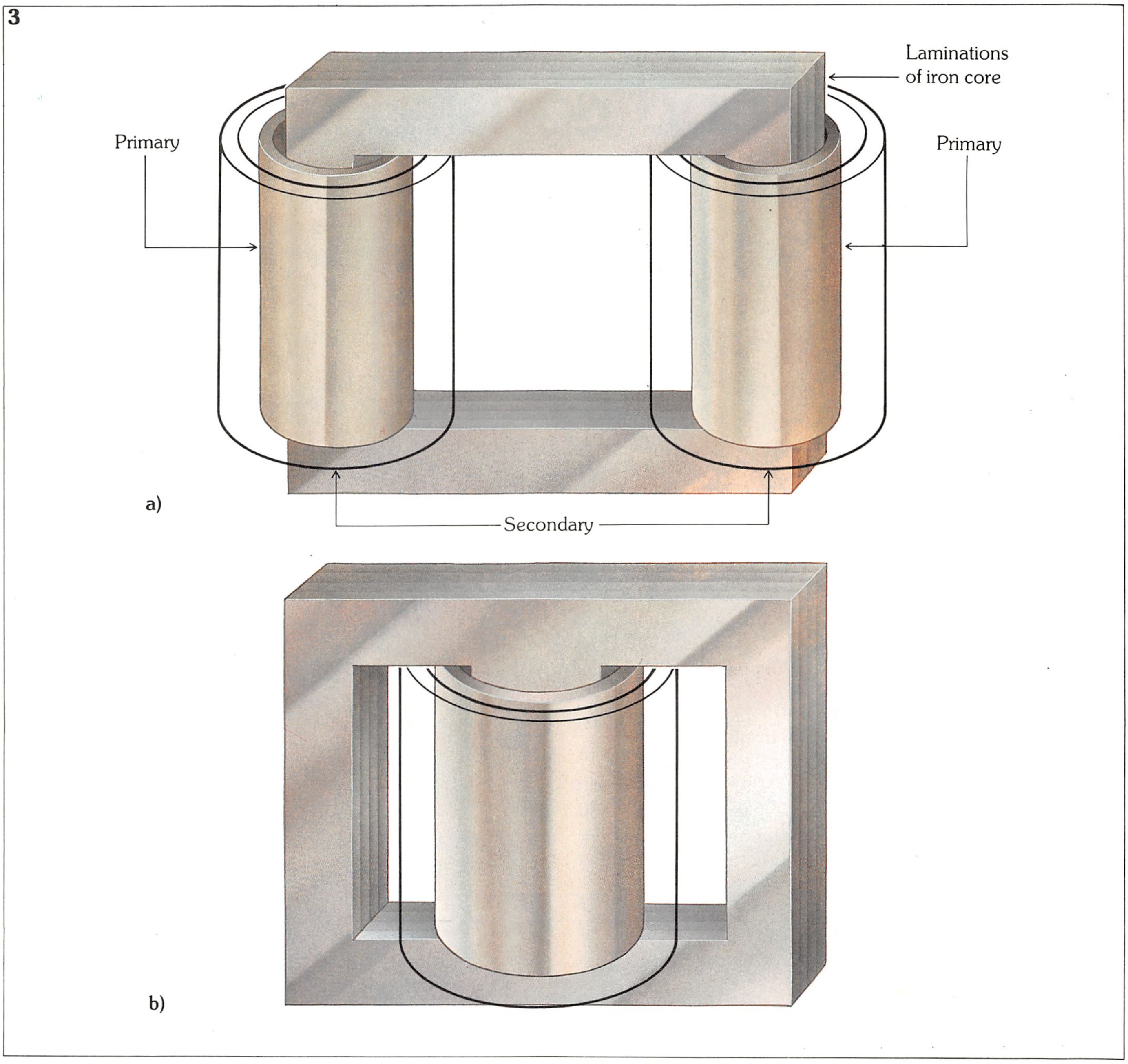
$$I_o = \sqrt{I_M^2 + I_C^2}$$

Since the terminal voltage changes very little as we transmit power through the transformer, the flux remains constant and so does I_M, I_C and I_o. So you can see that the no-load current is independent of the magnitude of the load current. The no-load current is of course measured when the transformer's secondary is not feeding anything.

Voltage regulation

Transformer coils are wound from copper wire, so we should expect to find a small **winding**

3



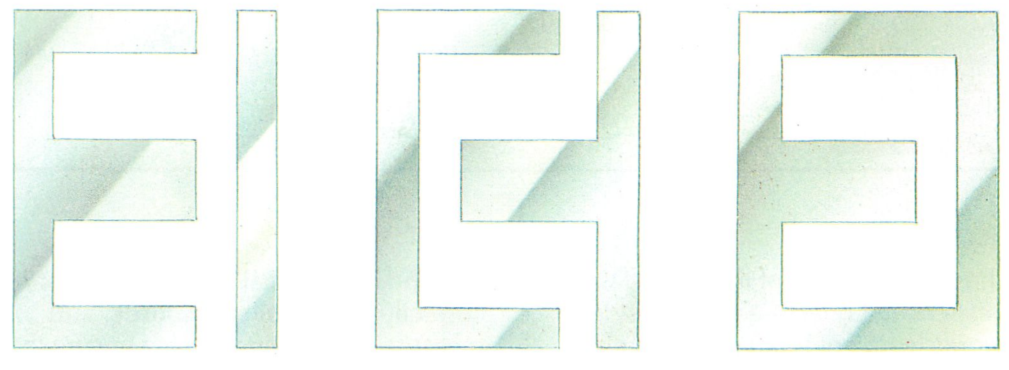
3. **Basic transformer constructions: (a)** core; **(b)** shell.

resistance in the primary and secondary windings, and as the load current flows through the secondary winding there will be a drop in voltage across it. As the load current increases, the output voltage no longer corresponds to the value obtained from the ratio of the turns on the two windings, but is slightly lower.

Not all the flux set up by primary current links with the secondary winding; a small amount of flux will link only with the primary. This represents a small self inductance in the primary circuit and a similar self inductance also exists in the secondary circuit. These reactances cause a fall in the output voltage below the open-circuit value, but this voltage fall is in *phase quadrature* with the load current. The effects of the winding resistance and leakage reactance may be represented by a resistance and an inductive reactance in series with either the primary or secondary.

The magnitude of this resistance and reactance may be obtained by short circuiting the secondary. The applied voltage will now only need to drive a current through the winding resistance and reactance, so if the magnitude of the current, voltage and phase angle were measured, the values of the resistance and reactance can be found. However, short circuiting a transformer's secondary winding would cause a very large current to flow,

4



4. Iron laminations in transformer core, which are usually made in two parts.

which might destroy the device. So, to overcome this, it is essential that the primary voltage is reduced, so that the short-circuit load current corresponds to the transformer's full-load current rating.

We have previously said that a transformer's output voltage falls as the current is increased. This is shown in *figure 2*; where graph a shows a typical transformer's variation of terminal voltage as current is increased for a resistive load. As you can see, the voltage falls about 10% from no load to full load. If an inductive load is used, the voltage variation is greater as shown in graph b. Finally, graph c shows that for a capacitive load, the output voltage rises as the load current is increased.

This variation of terminal voltage with load current can be expressed in terms of the transformer's **regulation**, which in the UK is defined as:

$$\text{regulation} = \frac{\text{no load voltage} - \text{full load voltage}}{\text{full load voltage}}$$

So we can see that regulation is positive for resistive and inductive loads and may become negative for capacitive loads.

Construction

There are two basic transformer constructions – core type and shell type, and these are shown in *figures 3a* and *3b* respectively. The advantage of the shell type construction is that the leakage flux can be made much smaller, thus improving the transformer's regulation. This can be further improved, if the coils are constructed in a sandwich. The construction shown in *figure 3b* has two coils wound one outside the other. If they are made in the form of a sandwich, with the primary windings at the top and bottom of the core's limb, and the secondary in the middle, the leakage flux can be further reduced. The drawings in *figure 3* are not representative of actual transformer designs as, in reality, the core is designed so that the windings completely fill the windows in the core.

We know that the core is made up of iron laminations, and these are generally made in two parts, as shown in *figure 4*. Successive laminations are inserted into the coils in opposite directions, so that the gaps alternate on either side of the transformer windings. The laminations are then bolted or clamped together, along with the hardware needed to mount the transformer. □

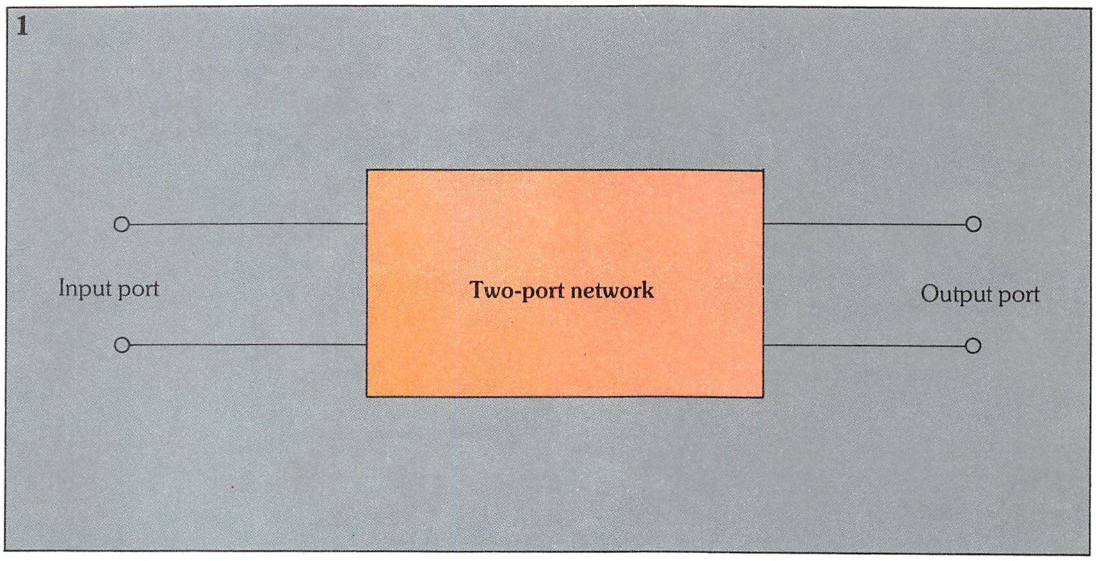
DIGITAL & SOLID STATE APPENDICES

6 Signal sources

We have seen in one of the *Basic Theory Refreshers* that often we may consider networks (some of them very complex networks) with one pair of input terminals and one pair of output terminals, as **two-port networks**. A two-port network may be considered as in *figure 1* as a block with an input port and an output port. It is not necessary that we understand the internal operation of the two-port network to consider what the network does. For example, the two-port network could be a radio receiver – the input port taking a radio signal from the aerial and the output port supplying audio signal current to a loudspeaker. Or, the network could be an audio high-fidelity amplifier – the input port accepting a signal from a record player cartridge and the output port supplying audio signal current to its loudspeaker. We can see that the operation of the network may be considered not as important as the input and output signals themselves.

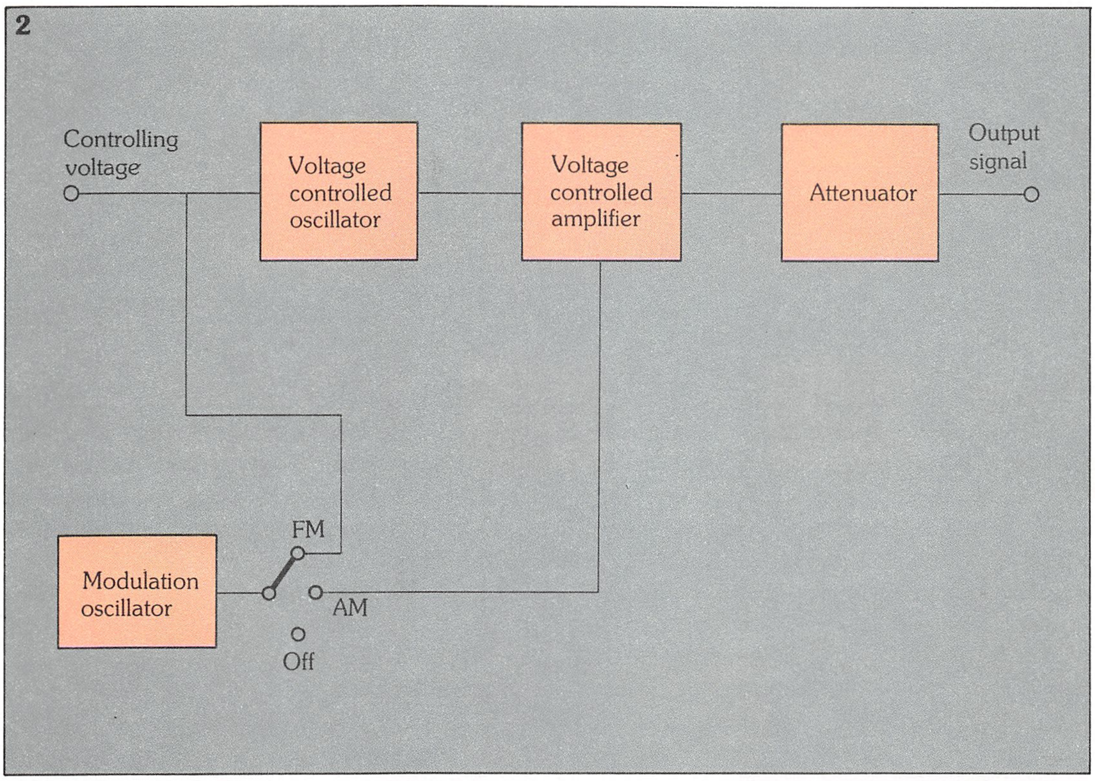
This block diagram, two-port network concept is useful when we wish to test the operation of a particular circuit or system. If, say, we provide the circuit or system with the input signal it requires to operate correctly, from a **signal source**, and measure the output signal to ensure the signal is within specifications, then we know the circuit or system is operating correctly.

The test equipment we have studied so far (analogue and digital multimeters, oscilloscopes) is designed to utilise this concept and measure a two-port network's output signal. In practical terms, of course, multimeters and oscilloscopes may be used to measure much more than just the output signal of a circuit or system, but nevertheless the concept is valid.



1. Two-port network.

The test equipment we shall look at now is designed to provide the necessary input signal(s) required by the two-port network under observation. As there are many different kinds of circuits and systems which may fit into the two-port network concept (analogue, digital, radio-frequency etc.) there are, inevitably, many different kinds of signal source test equipment. There are, however, general trends which such signal sources follow, and it is these trends we shall look at rather



2. A simple signal generator. It can simulate the signal from a received radio transmission.

than specific items of equipment.

Radio frequency oscillators

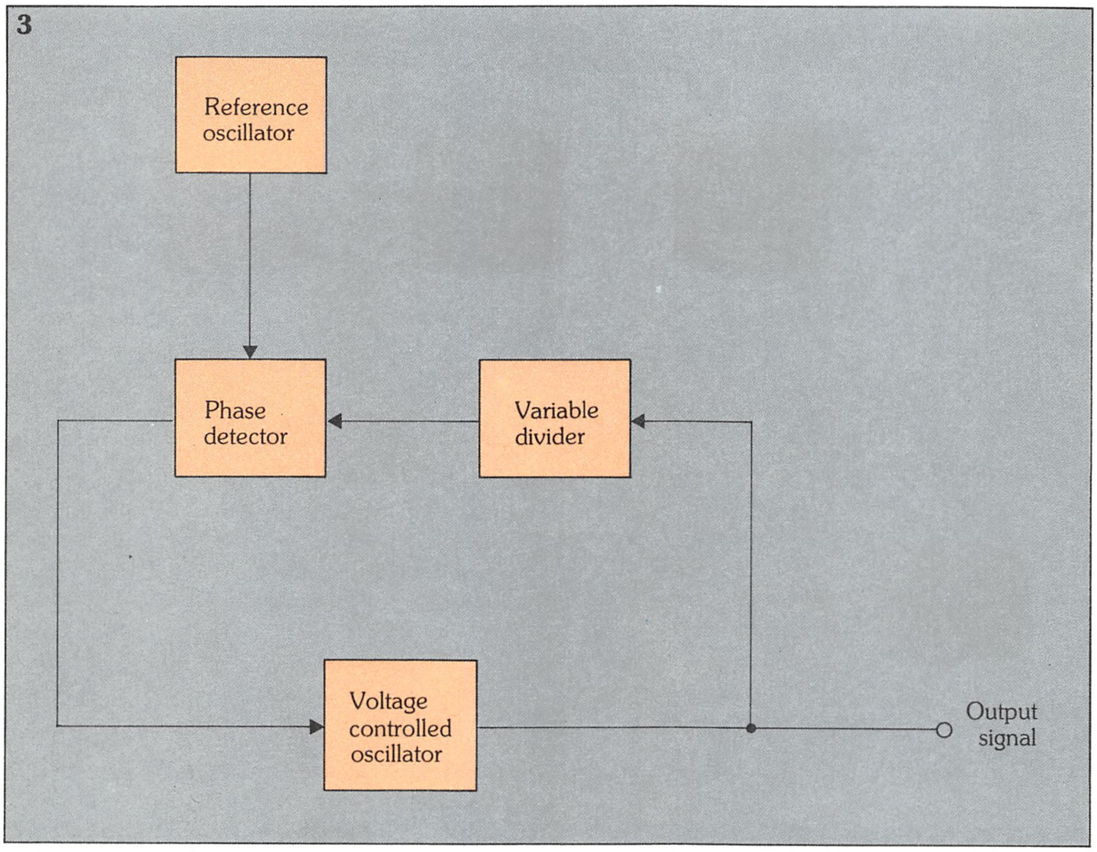
The first trend we shall consider comprises the signal sources used to provide radio frequency signals – **signal generators**. Such equipment generally provides a sine wave of radio frequency from an internal oscillator. It is usually possible to modulate the oscillator in a number of ways, e.g. amplitude modulation, frequency modulation, so that the signal generator may simulate to a reasonable accuracy the signal from a received radio transmission. Output signal from the generator is through a calibrated attenuator, so that a precisely known signal amplitude is available.

A simple signal generator with these facilities is shown in *figure 2*, in block diagram form. The oscillator used is a voltage controlled type (i.e its frequency is controlled by an applied DC voltage). Frequency modulation of the output signal can thus be undertaken simply by superimposing a small AC signal onto the DC controlling voltage. Generally, such oscillators are formed around an LC tuned circuit and different values of coil or capacitor are switched in and out of circuit to give various ranges. The voltage control facility may be included by the use of a **varicap diode** (an abbreviation of *variable capacitance*), a semiconductor device which exhibits a junction capacitance which varies, depending on the voltage across the junction.

Amplitude modulation of the oscillator output signal is undertaken with a voltage controlled amplifier. The modulation oscillator is generally a low frequency sine wave oscillator and may or may not be variable in frequency. Often, the internal modulation oscillator can be switched out of circuit and an external signal applied to modulate the radio frequency signal. If, say, a music signal or a voice signal is used to modulate the radio frequency signal, the modulated output signal would, therefore, more accurately represent a possible received radio transmission than does the sine wave modulated signal.

The attenuator is a simple potential divider network, with variable resistors, allowing a proportion of the oscillator's output signal to be available at the signal generator's output.

Frequency of oscillation of this type of signal generator is not particularly stable



3. Synthesiser signal generator which overcomes the problem of instability of frequency of oscillation found in simple signal generators.

and, in fact, may often drift greatly from the required frequency. One signal generator method overcomes this problem in what is known as the **synthesiser** principle, shown in *figure 3*.

A synthesiser signal generator uses a fixed frequency reference oscillator which, because it *is* of fixed frequency, may be very stable. A phase detecting circuit compares the phase of this reference oscillation with the phase of the signal from a voltage controlled oscillator after it has first passed through a variable divider circuit.

The output of the phase detecting circuit provides the controlling voltage for the voltage controlled oscillator.

If the reference frequency oscillator signal leads the divider circuit output signal, the frequency of the divider signal is lower than the reference, and so the output voltage of the phase detecting circuit changes to adjust the voltage controlled oscillator. If, on the other hand, the reference signal lags the divider signal, the divider signal is higher in frequency than the reference signal and the output voltage of the phase detector circuit again changes (the other way, this time). A feedback loop is therefore generated which tends to stabilise the frequency of the voltage controlled oscillator at a frequency equal to the reference oscillator frequency times the division ratio of the divider circuit.

As this synthesiser system uses a feedback loop, and as the output signal frequency is said to be **phase-locked**, the principle is sometimes known as a **phase-locked loop** oscillator. The phase-locked loop principle has many other applications, for example in modern synthesised radio receivers – the type with a digital display of received frequency.

Often the division ratio of the divider circuit is microprocessor controlled and so the user simply enters on a keyboard the desired signal frequency. The microprocessor defines the divider circuit's division ratio and the oscillator automatically outputs a signal of the required frequency.

In such **digitally synthesised** signal generators the remainder of the circuit may be made up with the individual blocks as in *figure 2*, and indeed, they too may be microprocessor controlled so that the user

sets required attenuation, depth of modulation etc., via the keyboard.

Low frequency oscillators

Low frequency oscillators are used to simulate the input signals to low frequency systems. Such systems include audio systems: high fidelity sound, telephone communications, public address etc., but we must also include those systems which require extremely low frequency signals – some control systems for example – or those which require signals of frequencies above the audio range – ultrasonic remote control for television receivers etc. The frequency range of low frequency oscillators must extend from, say, 0.01 Hz to 1 MHz.

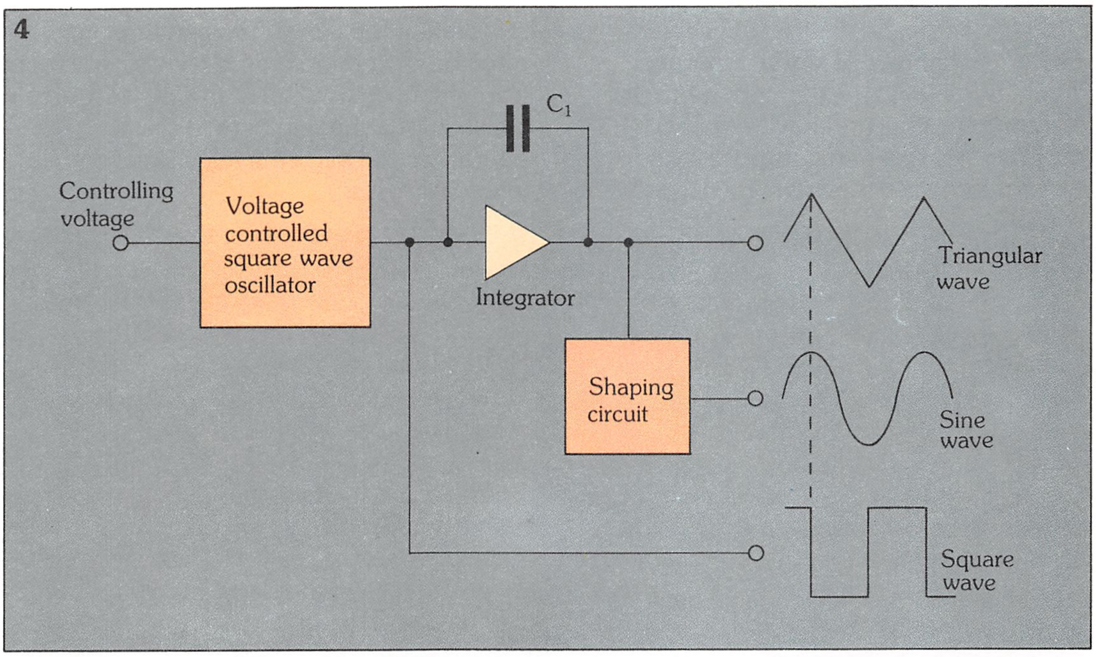
Just as there are many types of radio frequency oscillators, there are a number of different low frequency oscillator principles. The simplest is the type based on resistor-capacitor circuits such as the Wien bridge oscillator or the phase shift oscillator. We saw these in detail in *Solid State Electronics 27* so we need not discuss them more than just in passing here.

A slightly different principle is used in the type of oscillator known as the **function generator**. A block diagram representing its basic operation is shown in *figure 4*. The main block of the function generator is an integrator which produces a triangular waveform from an applied squarewave waveform. A voltage controlled oscillator producing a squarewave output is used to generate the signal at the required frequency. In a practical circuit the oscillator and integrator are not, in fact, entirely separate blocks as they are more interdependent than the block diagram shows. For example, the value of the integrator capacitor C_1 determines the basic frequency range within which the function generator operates, so switching different values of capacitor in and out of circuit allows different ranges to be covered. The function generator's voltage control input allows fine tuning of the frequency within the chosen range.

A range of output waveforms is available from a function generator. By applying the triangular wave signal to a shaping network the result is a type of sine wave output signal. So, triangular wave, square wave and pseudo-sine wave output signals are available.

Digital low frequency oscillators

It is possible to store the digital values corresponding to a waveform in a ROM device. This may be undertaken by sampling the waveform, say, a sine wave, making an analogue-to-digital conversion of each sampled value, and then programming the ROM device with the



4. Function generator which can give a range of output waveforms.

resultant digital codewords. An oscillator may then be formed simply by reading, in sequence, the digital codewords stored and making a digital-to-analogue conversion of them.

Frequency of oscillation is determined purely by the rate at which the ROM is read. Varying the reading rate varies the output signal frequency.

Output waveform quality depends primarily on the accuracy of the stored digital codewords but also, of course, on the resolution of the system which is affected by the number of bits in each stored codeword; 8-bit codewords, for example, give a resolution of 1 in 255 (i.e. 1 in 2^8-1).

Other categories of signal sources

These two main types of signal sources: signal generators and low frequency oscillators, are general purpose items of test equipment – they may be used to provide input signals for the majority of electronic circuits and systems.

However, as more and more complex electronic systems are developed, in particular digital systems, it becomes necessary to produce more and more specific signal sources. For testing logic circuits, for example, **pulse generators** are often used as a source of defined input pulses. Such pulse generators normally allow the user to define pulse amplitude duration, and frequency.

Television receivers, too, may often be tested with the use of the specific signal sources – **television test signal generators** – which produce signals corresponding to full television signals received at the aerial. Test patterns may often be generated on the television screen from the output signal of the signal source, which allows the engineer to check operation and align the television receiver where necessary.

Glossary

digital synthesis oscillator	oscillator using the phase-locked loop principle, whose divider circuit's division ratio is microprocessor controlled. The user may thus determine the output signal frequency by entering information via a keyboard
function generator	a low frequency signal source which utilises an integrator to convert a square wave signal into a triangular wave. An approximation to a sine wave is given by shaping the triangular wave
phase-locked loop	a synthesis oscillator system in which the output signal is phase locked to a reference oscillator signal. Frequency of the output signal is defined by the division ratio of a divider circuit
pulse generator	signal source which produces pulses of user-defined amplitude, width and frequency
signal generator	signal source of radio frequency oscillation signals
synthesised oscillator	see phase-locked loop
varicap diode	a diode which displays a junction capacitance. The capacitance varies in value depending on the voltage across the junction

COMMUNICATIONS

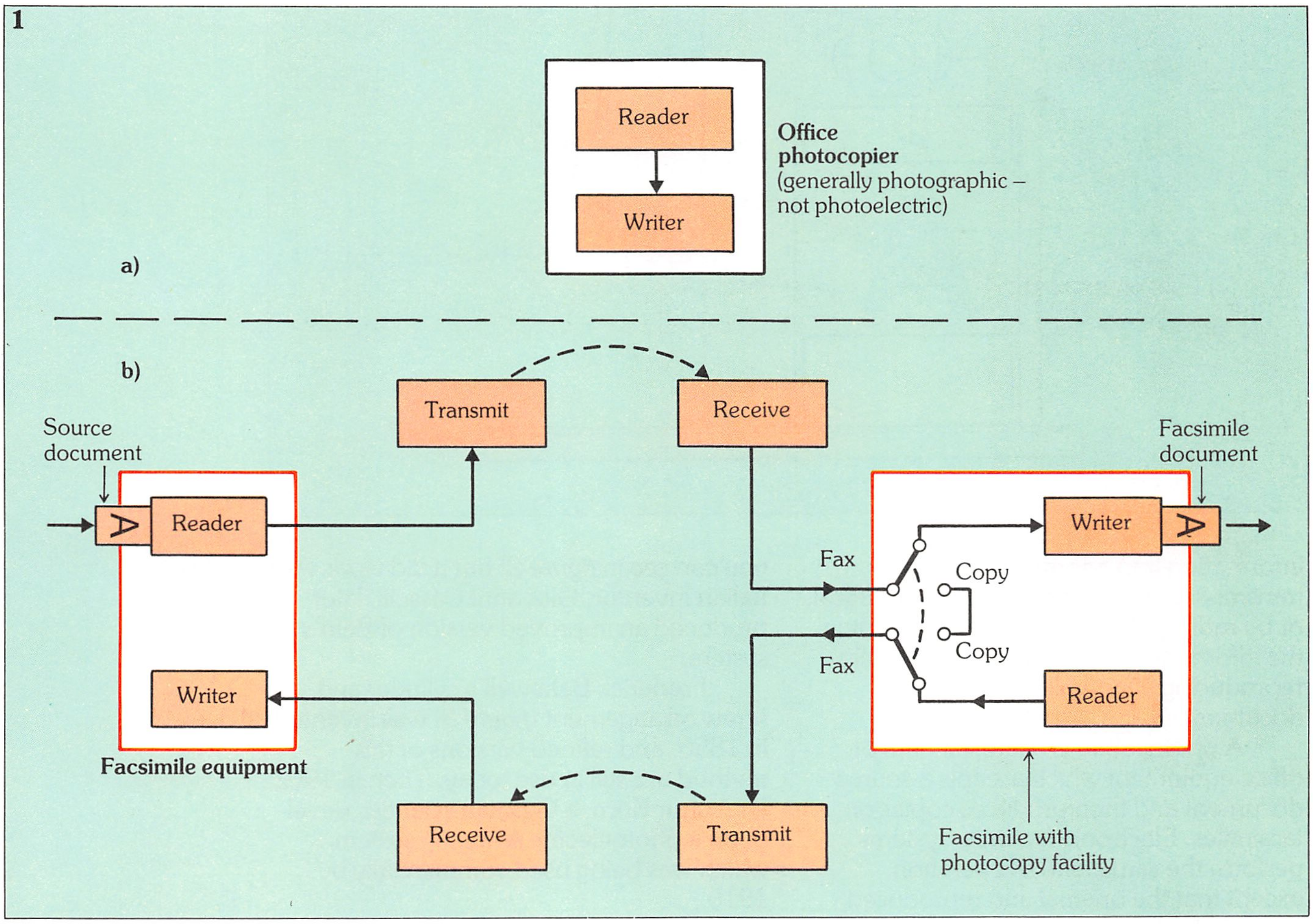
Facsimile systems-1

What is facsimile?

So far in our examination of communications systems, we've looked at the transmission of sound and visual information (via the telephone system, television and radio) and also briefly mentioned the transfer of digital computer output across existing telephone networks. These communications methods allow a permanent copy of the transmitted information to be made at the receiving terminal: sound signals can be recorded onto magnetic tape: typed information, either a telex message or computer data input on a keyboard, can be reproduced as a 'new' typewritten document, or stored permanently using another medium such as magnetic tape or floppy disk; and television signals can be stored on video tape.

But what about a drawing, a map, photograph or fingerprint? Such information cannot be typed into a keyboard, or described adequately over the telephone, and televising it would be far too cumbersome and expensive. For such information, a communications method known as **facsimile** is used.

A facsimile is an exact copy or replica of an original or **source document**. (For



1. Office photocopier compared with an electronic fascimile system.

example, a 'facsimile edition' is an exact reproduction of an original and, usually rare, book or document. Limited facsimile editions of such things as the first printed bible and the Magna Carta are sometimes published at considerable cost, and purchased by libraries and collectors.)

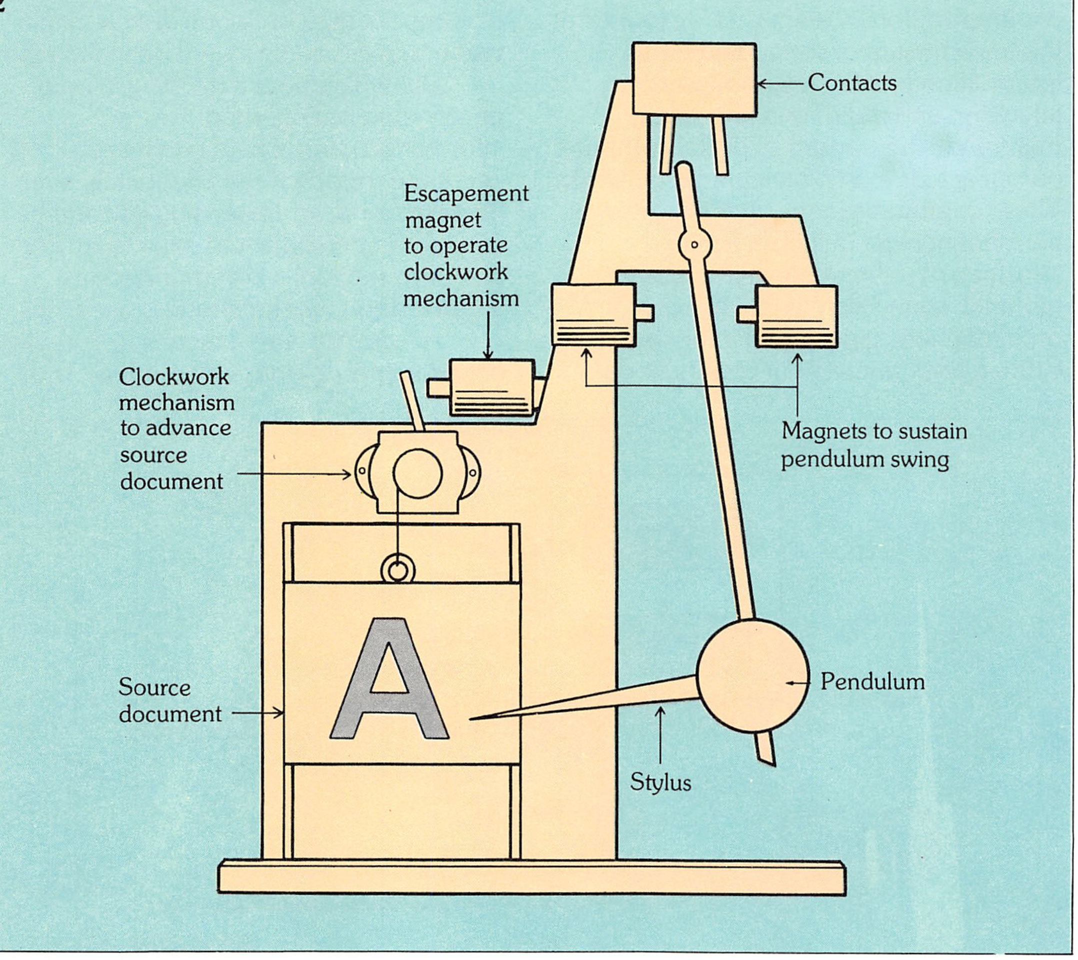
In the context of electronic communications, the term facsimile system refers to the techniques used in copying a source document by encoding the information into a form suitable for transmission (over cable, coaxial or optical, or by radio networks); and then decoding the information at the receiver, faithfully reproducing a facsimile of the source document.

A photocopier is a familiar item of office equipment which accepts a source document and then produces copies or facsimiles. Electronic facsimile systems perform the same kind of operation, except that the original and reproduced documents may be separated by hundreds or even thousands of miles and the method of reproduction is different. These differences are illustrated in *figure 1*.

You might think that facsimile (or **fax**) is only a very recent development, however, it is almost as old as the telegraph! The first recognisable facsimile system was developed by a Scotsman, Alexander Bain, in 1842 and patented by him in 1843. Obviously it was clumsy (as you can see in *figure 2*) but it did work. An Italian inventor, Giovanni Casselli, later produced an improved version of Bain's system.

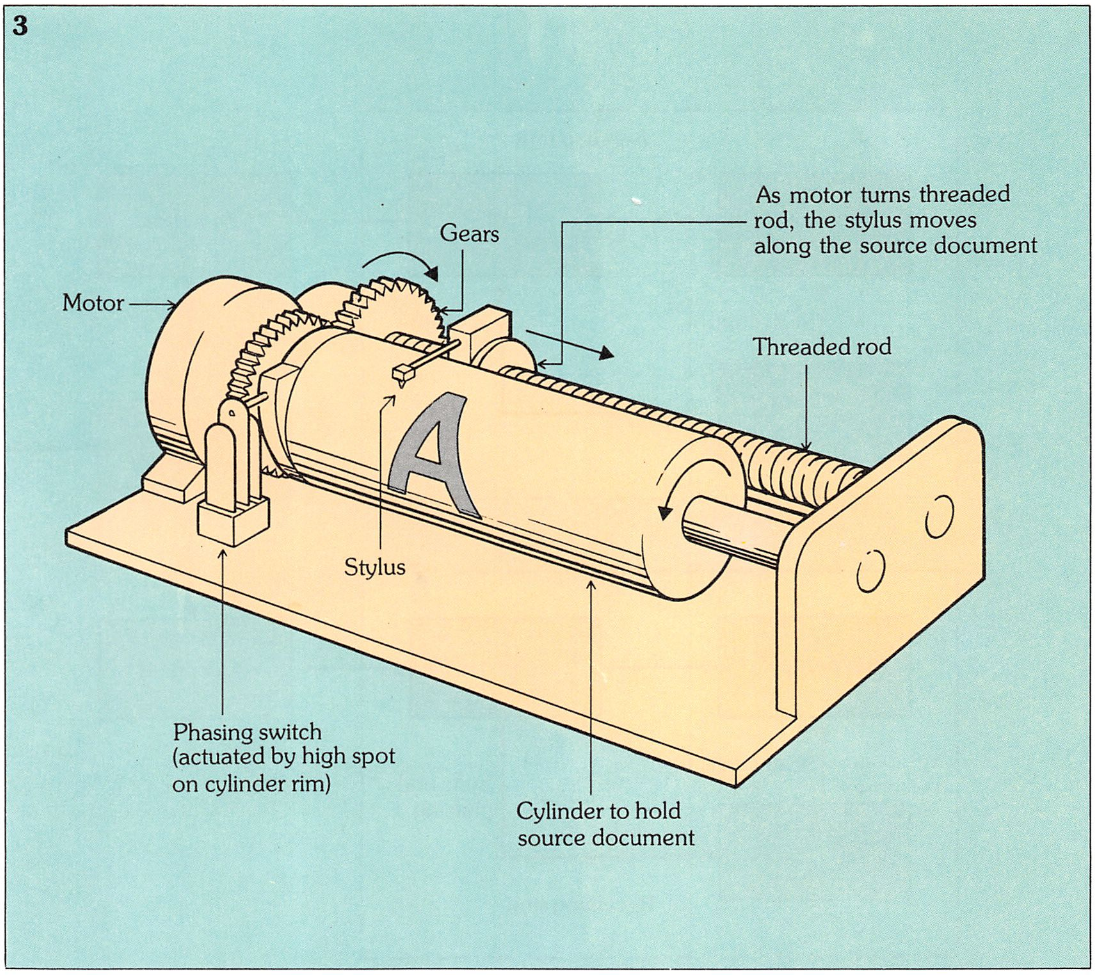
Frederick Bakewell's cylinder and screw arrangement (*figure 3*) was invented in 1850, and refined versions of this method are still in use today. Then in 1902 Dr Arthur Korn, a German scientist, developed a photoelectric scanning system which was being used commercially by 1910.

2



2. Alexander Bain's facsimile system, patented in 1843.

3. Frederick Bakewell's invention in 1850 – refined versions of this cylinder and screw arrangement are still in use.



By the 1930s, facsimile was established as a useful communications medium and was being used for a variety of purposes, many of which have endured until the present. One idea which didn't last was the use of facsimile to transmit complete newspaper pages to receivers installed in subscribers' homes. This interesting and original idea predated teletext, its modern equivalent, by nearly 50 years. Unfortunately for the speculators who conceived the plan, World War II intervened – after the war, they could not compete with the instant and overwhelming attraction of television.

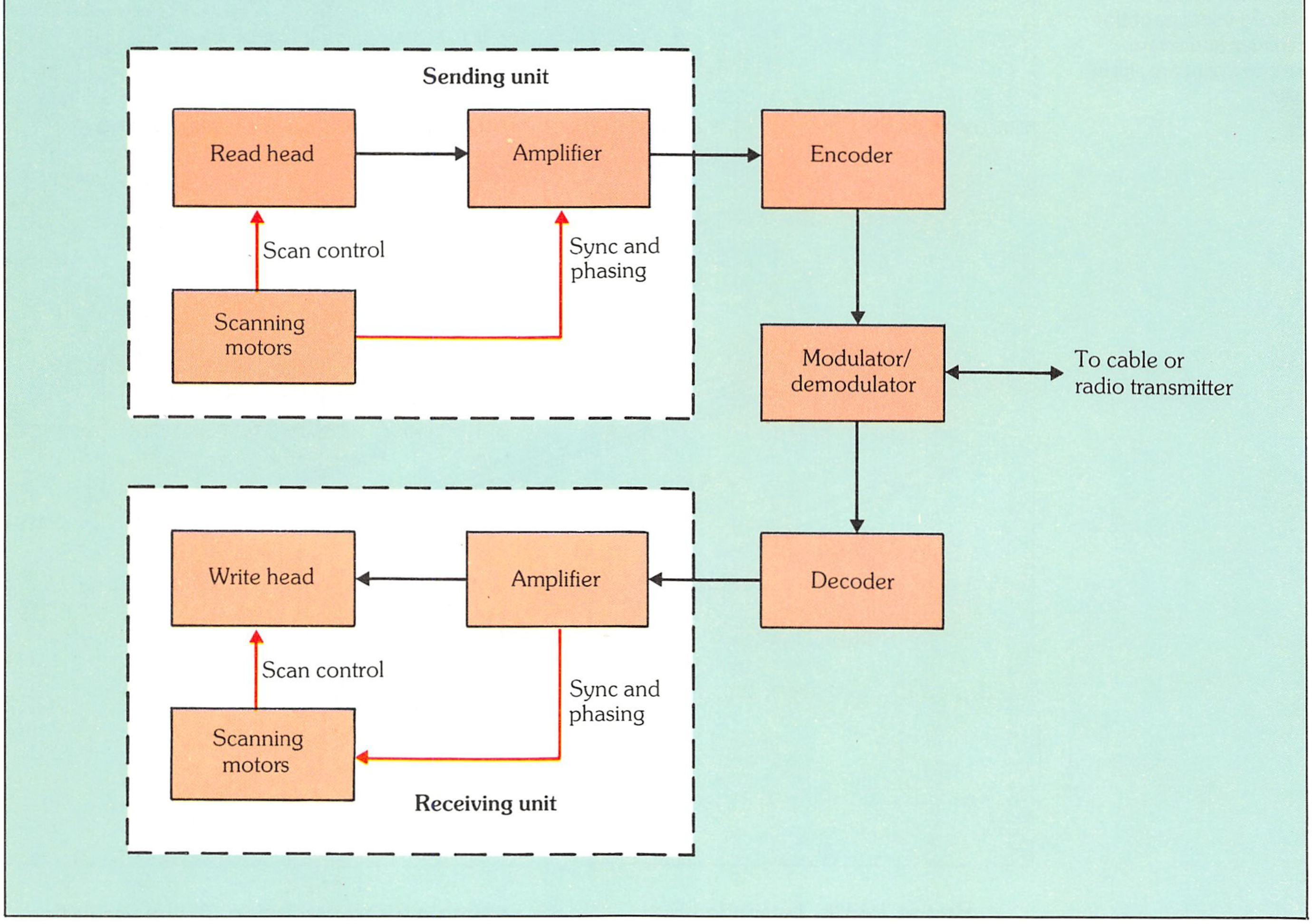
The elements of a facsimile system are shown in block diagram form in *figure 4*. The source document is placed in the sending unit, where the visual information is converted into an electrical signal. This signal is an exact representation of the visual information, an **analogue** of the source document. The electrical signal is then converted by an encoder into a form suitable for transmission. At the receiver, this process is reversed. First, the information is extracted from the radio or telegraph signal, and then it is decoded to produce a signal the same as that first produced by the sending unit. This signal is known as the **baseband signal**, and at the receiver it is used to control a printing device.

It is worth noting, at this point, that because the receiving unit must be able to produce a document substantially similar to the source document, certain characteristics of the sending and receiving units must be identical. This problem of **standards** will be discussed in detail in *Communications 6*.

Applications

Facsimile offers a number of advantages over other communications systems: it is flexible, because it can deal with any printed or photographic image; and it is forgiving, because small errors in

4



4. Elements of a facsimile system.

transmission do not destroy the **integrity**, i.e. the wholeness of the image. The only disadvantage of facsimile is that it can be expensive when telephone charges are high. However, it is the only method available for the rapid communication of printed, photographic or drawn images over large distances.

One of the first uses for facsimile was the transmission of meteorological information, and today, complete weather charts are sent between central and local meteorogolical offices, to TV stations, newspapers, ports, airports and to ships at sea. Consider the amount of information contained in a typical weather chart – geographical features, isobars, wind speed and direction, temperature and so on. This data could be tabulated and transmitted as a numerical list, however, the numbers would then have to be plotted onto an appropriate map – a lengthy operation.

Newspaper pages are routinely sent by facsimile between publisher and printer. For example, the European edition of the Wall Street Journal is printed on the Continent from whole pages sent by facsimile from the U.S. In addition, since the European edition can be printed directly from the facsimile, using the output document as the printing master, there is no duplication of labour.

One very common use for facsimile is in the worldwide distribution of news photographs to agencies and subscribers. (The potential of facsimile for this purpose was realised early in its history, when a photograph of Pope Pius XI was sent from Rome to the U.S. and published in a New York paper on 11th June 1922, the same day it was sent.) Other applications include the transmission of fingerprints, police records, engineering plans and drawings and signature verification.

Facsimile in operation

Returning to *figure 4*, we see that the read head or detector scans the image (the source document) and creates an electrical analogue of that image. This analogue signal is then encoded into a form more suitable for transmission, via a modem, over a standard communications link (this stage is omitted in some systems) such as the public telephone network.

A modem (**mod**ulator/**dem**odulator) is a device which converts signal output from one piece of equipment into a form suitable for transmission along a communications link for input to another piece of equipment. It achieves this by modulating and then demodulating a carrier wave in order to represent data on a telephone network. (Modems will be further discussed in greater detail in *Communications 10*.)

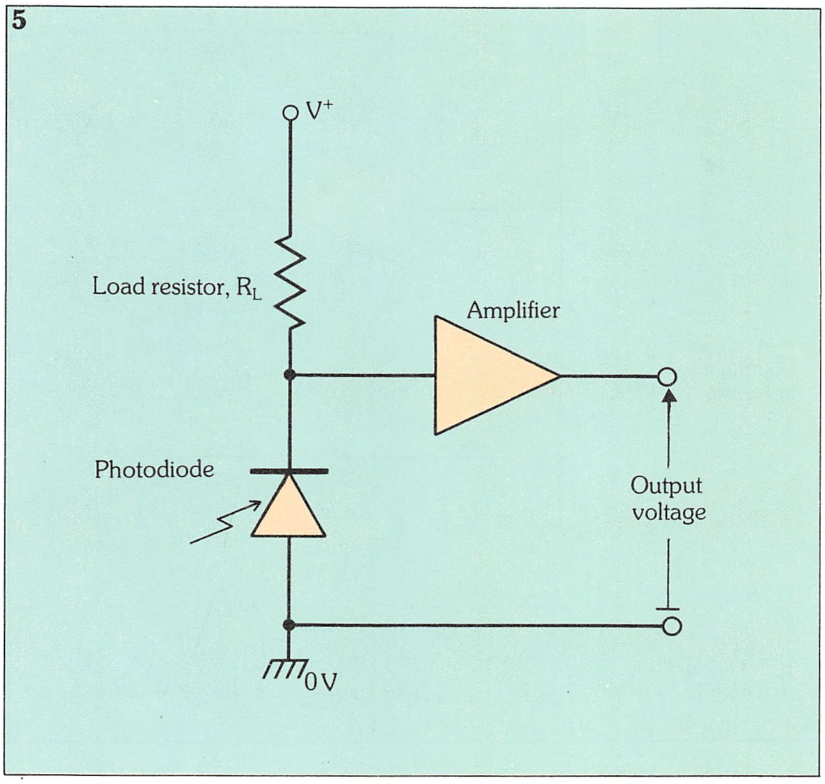
The telephone network, as we have seen, comprises fairly low grade communications channels, and the limitations of telephone systems impose severe restrictions on what can be achieved in facsimile transmissions, as we will see in the following article, *Communications 6*.

At the receiver (*figure 4*), this process is reversed. The modulated signal is demodulated, decoded (if necessary) and thus converted back into a baseband signal (a replica of that produced by the read head). This signal then controls a reproduction device to produce the hard-copy paper output.

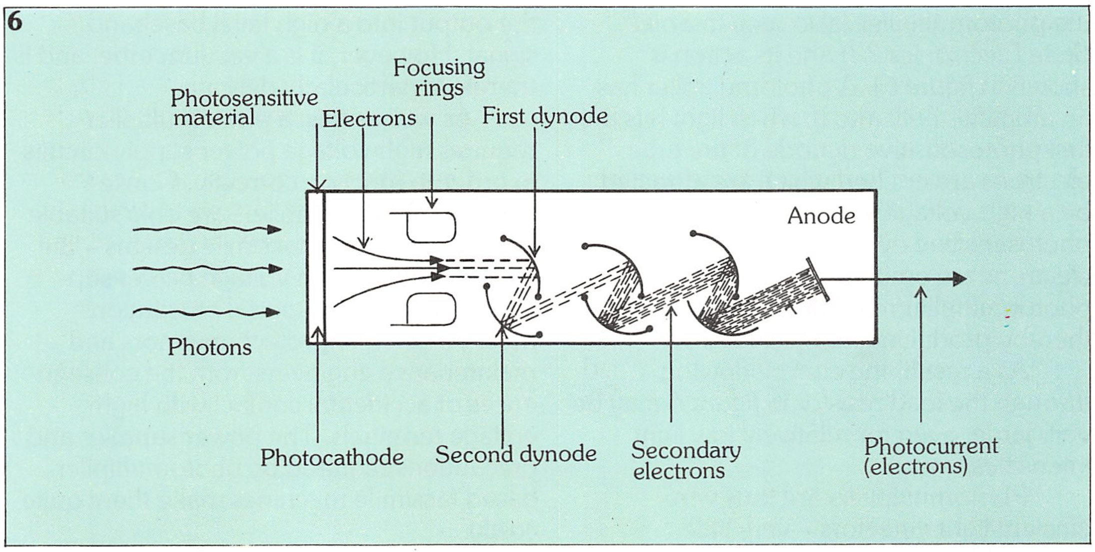
Reading the image

As we have said, facsimile transmissions begin by 'reading' the source document, that is, by sensing the printed, drawn or photographic information. In general, this is achieved by dividing up the document into small areas – known as **picture elements**, or **pixels**. The document is then illuminated, and the reflected light from each pixel is recorded. A scanning

5. A typical photodiode circuit.



6. Action of a photomultiplier.



mechanism sweeps across the width of the document (either smoothly or in very small stepped increments) from left to right and steps down, one line at a time, so that the whole image area of the document is sensed.

The varying light levels reflected from (or in some cases, transmitted through) the document represent the different densities of ink and hence the different shapes of letters, lines etc. on the document and can be used to vary the amplitude of a voltage or current, thereby producing a signal which is the electrical analogue of the image.

Several devices are available to perform this conversion, and as many different techniques are used to carry out the process.

The photodiode (previously discussed in *Solid State Electronics 22*) is one device frequently used to record the fluctuating light levels reflected from a source document. A typical circuit is shown in *figure 5*. As we know, a photodiode can be thought of as a semiconductor device which produces a current when light strikes the sensitive surface of a photodiode. The brighter the light falling on the photosensitive surface, the larger the diode current. By connecting the photodiode so that this current flows through a resistor, as in *figure 5*, the current is converted into a voltage proportional to the light falling on the photodiode. This voltage can then be amplified, and in a facsimile system this amplifier output is the baseband signal.

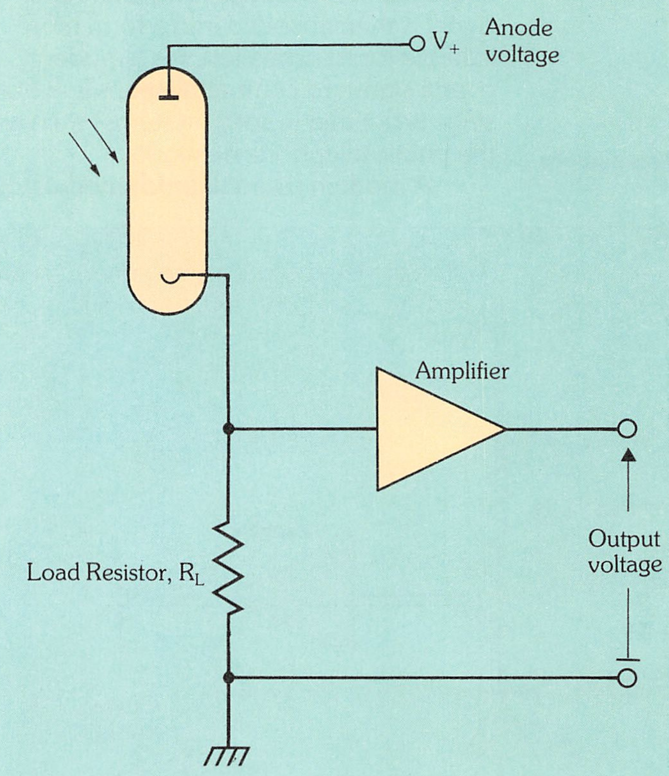
A second commonly used device is the photomultiplier (also seen in *Solid State Electronics 22*) and its action is shown in *figure 6*). A photomultiplier has an amplifier built into it: when light hits the first photosensitive dynode of the tube, electrons are emitted which are attracted by a high voltage towards a second photosensitive dynode, where more electrons are emitted, and so on. A photomultiplier may have many dynodes, thereby producing many electrons.

As a result, the current flowing through the load resistor in *figure 6* may be very large, even for relatively low light intensities.

Photomultipliers are thus very efficient light detectors – very little additional circuitry is required to convert the output into a high level baseband signal. However, it is a vacuum tube, and is therefore particularly delicate.

As well as this, a photomultiplier requires high voltage power supply circuits in order to function correctly. Consequently, photomultipliers are only suitable for certain types of facsimile designs – the requirement of high voltage power supplies meaning that special precautions must be taken to protect operators and maintenance engineers from the consequences of accidental contact with high voltage terminals. The power supplies and precautions required by photomultiplier-based facsimile machines make them quite costly.

7



7. Current flowing through the load resistor in a photomultiplier.

Scanning the source document

It is fairly easy, then, to produce an electrical signal proportional to a light intensity, but how is the image scanned so that information from every part of the source document is converted into an electrical signal? It must be obvious that this scanning method must be very precise in order to reproduce an *exact* copy of the original. Bakewell's cylinder and screw technique is one method, and although very fast electronic scanning methods are replacing these complicated mechanical devices, for some purposes (as we shall see), mechanisms very similar to Bakewell's are still in use today.

As we have said, the source document is divided up into pixels, if this were not done then a scan of the image would produce just one reading proportional to the average light intensity reflected from the whole page, and would not, of course, represent sufficient information for the source document to be recreated.

There are two methods of illuminating the source document. The first, **spotlighting**, illuminates each pixel separately – the reflection being collected and focused onto the photodetector; the second method, **floodlighting**, illuminates the entire image area at the same time, but the reflected light is limited to each pixel.

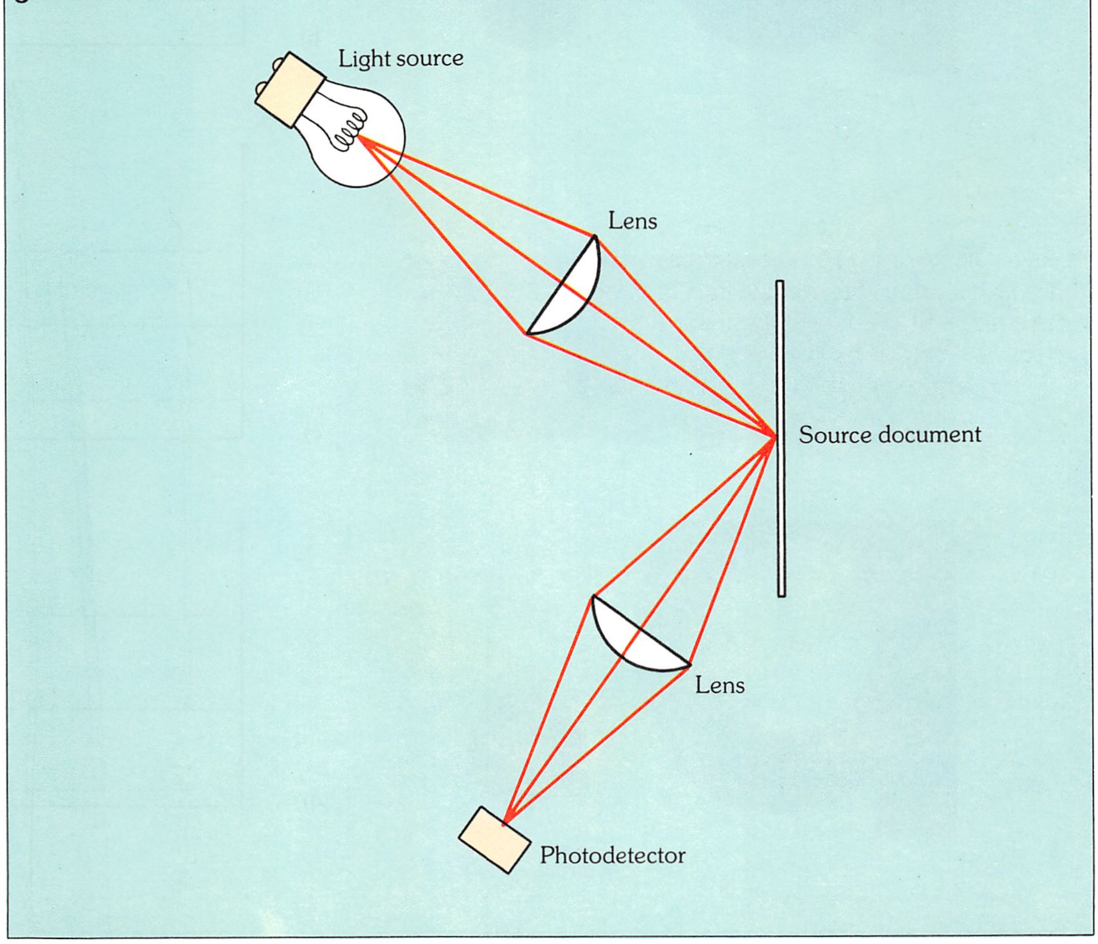
We will now consider each method in more detail.

Spotlighting

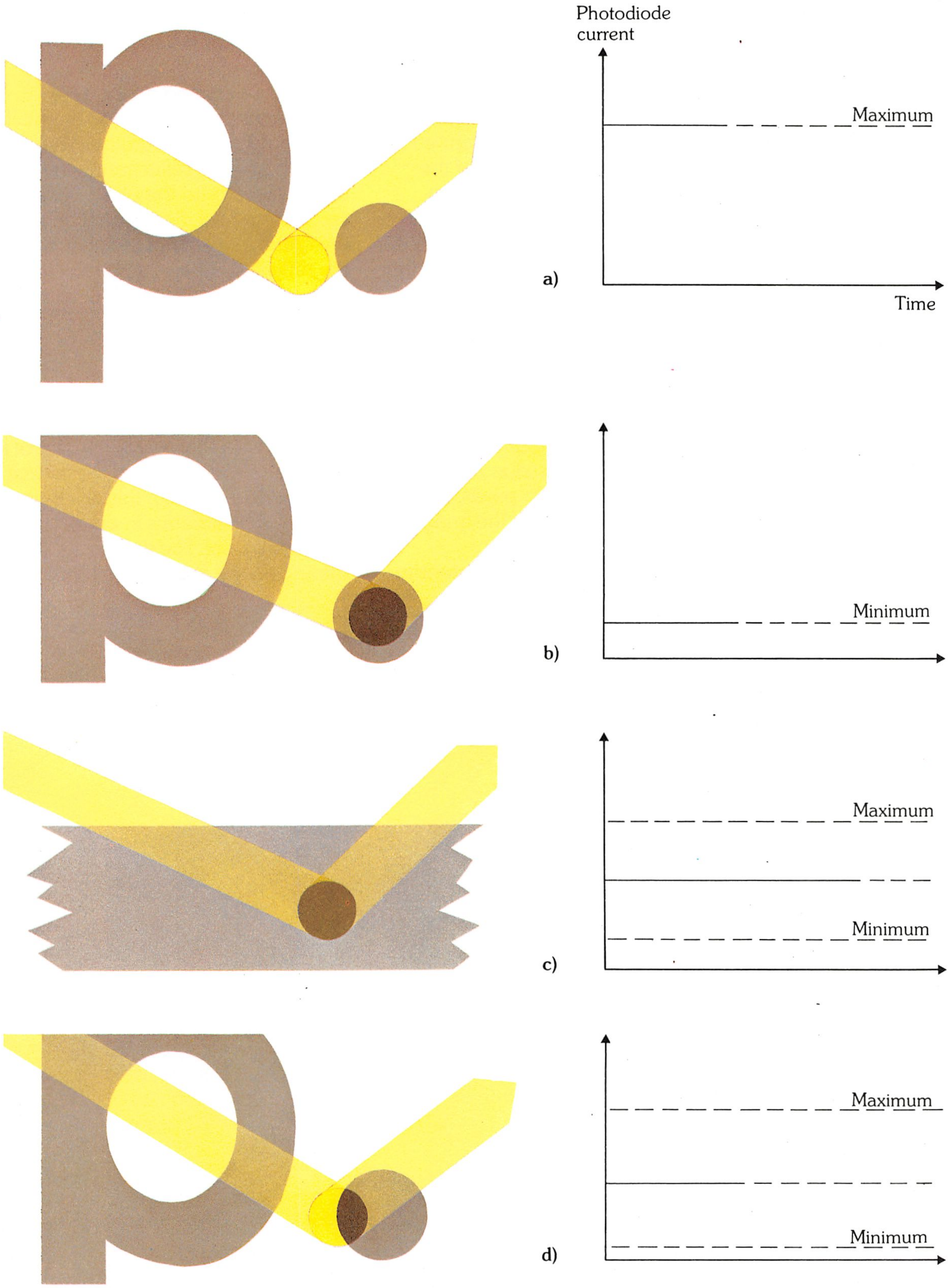
Spotlighting is illustrated in *figure 8*. The 'spot' is formed by focusing a light beam with a lens so that a very tiny dot of light illuminates the surface of the source document. Typically, this dot is smaller than 0.25 millimetres in diameter – smaller than a full stop. The tiny dot of light shining on the page causes light to be

8. Spotlighting.

8



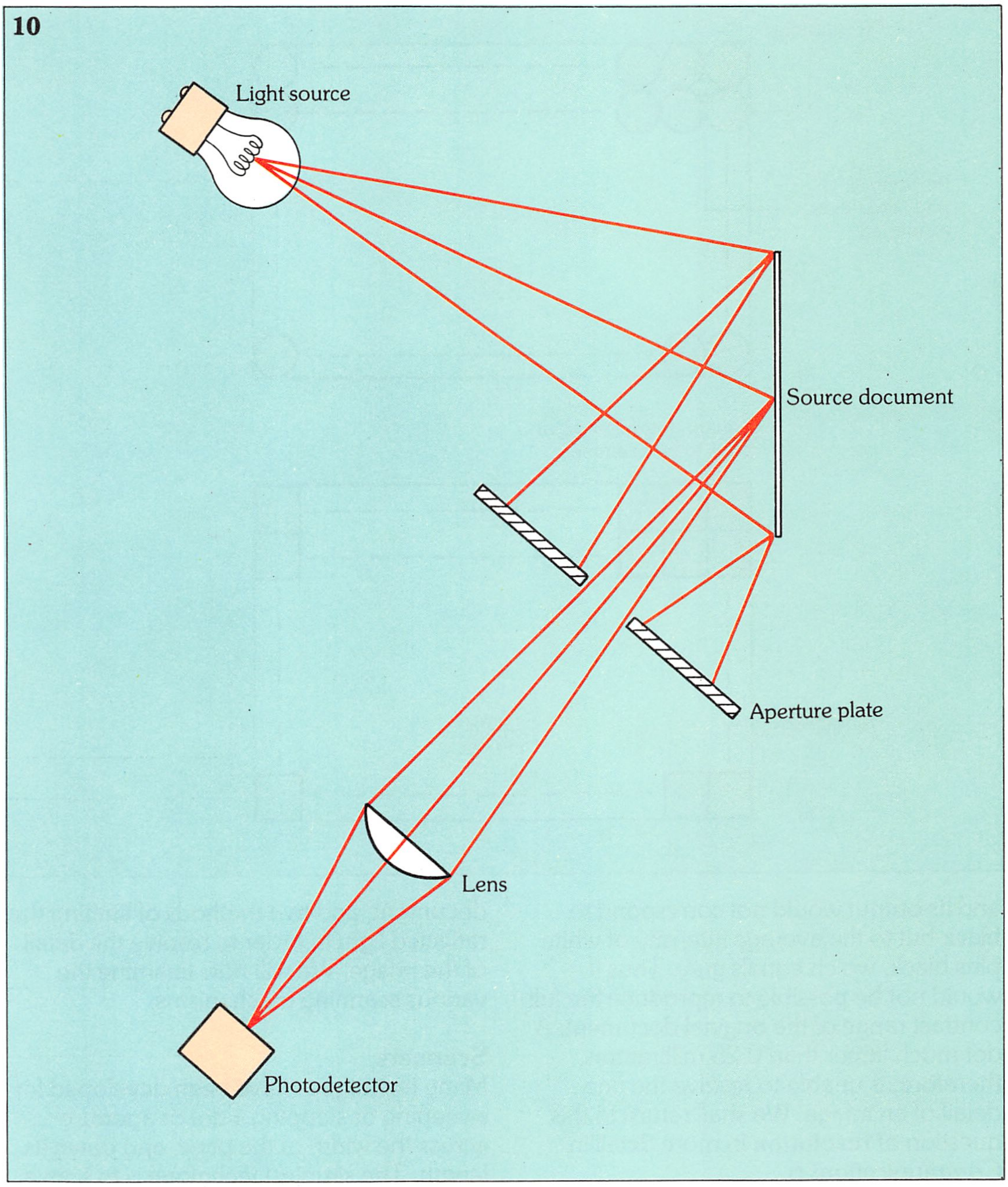
9



Photodiode current
Maximum
a)
Time
Minimum
b)
Maximum
Minimum
c)
Maximum
Minimum
d)

9. **(a)** If the illuminated spot is pure white, a maximum signal from the photodetector will be produced; **(b)** if the spot shines on a dense black area, the output signal will be minimum; **(c)** shades of grey in the image cause the photodetector output to fall between maximum and minimum; **(d)** photodetector output reacts to the average intensity of the reflected light.

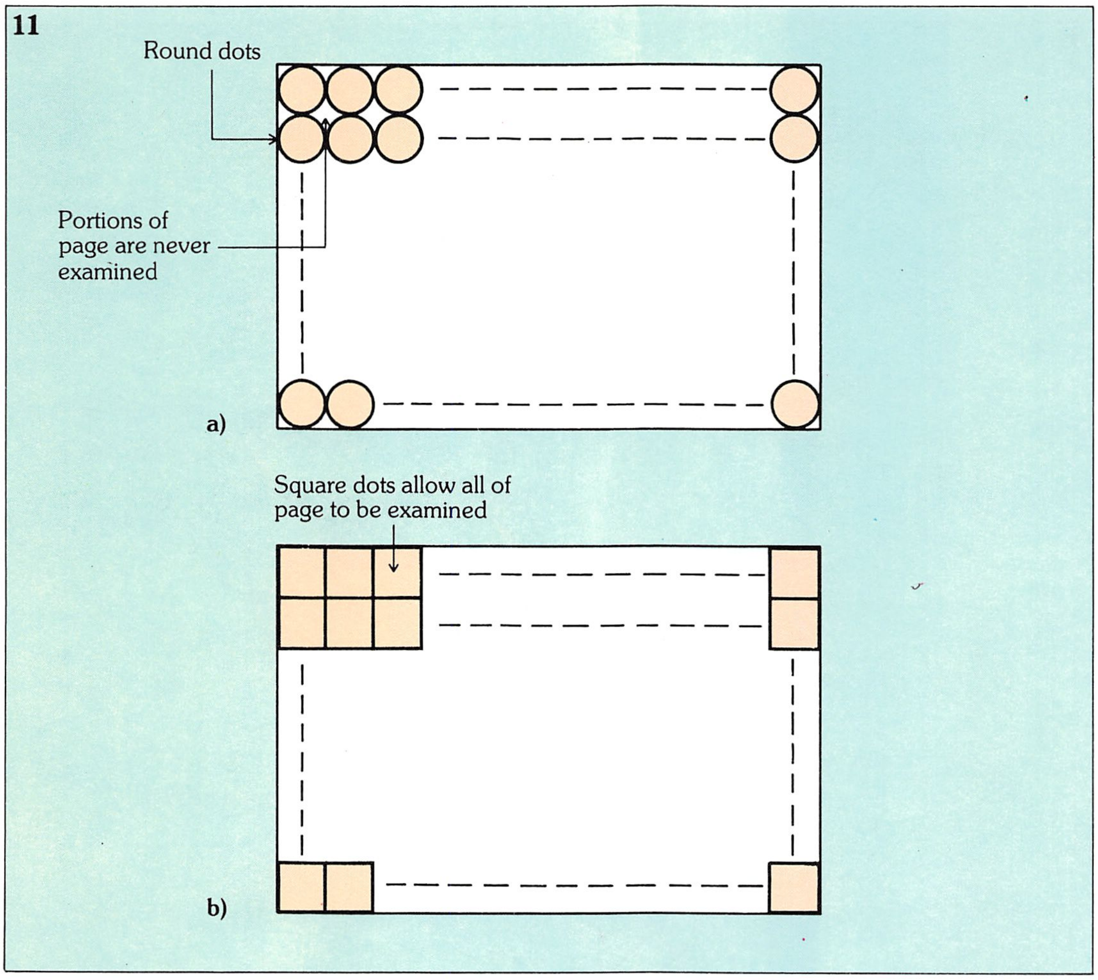
10. Floodlighting



reflected from the image, and the intensity of the reflection is proportional to the lightness or darkness of the pixel that is illuminated. The reflection is then collected by a lens system and focused onto the active surface of a photodetector.

If the illuminated spot is pure white, there is strong reflection producing a maximum signal from the photodetector (*figure 9a*); if the spot is shining on a dense black area, then very little light is reflected to the photodetector, and the output signal will be minimum (*figure 9b*). Any shades of grey in the image will, of course, reflect an intermediate level of light proportional to the density of the grey, and the photodetector output therefore falls between maximum and minimum (*figure 9c*). Similarly, if the spot shines on an area that is partly pure white and partly solid black, the photodetector output again reacts to the average intensity of the reflected light (*figure 9d*).

It should now be obvious why the dot must be very small: if it were slightly bigger than a full stop, for example, the photodetector would 'see' not just the black, but some of the white surrounding it,



11. Raster scanning: (a) surface of the source document broken down into circular dots; **(b)** into square dots.

and its output would not correspond to black but to the average intensity of white plus black, which equals grey. Thus it would not be possible to reproduce the full contrast range of the original document. A dot much larger than 0.25 millimetres, therefore, is unable to resolve the fine detail of an image. We shall return to this question of **resolution** in more detail in *Communications 6*.

Floodlighting

The principle behind floodlighting is shown in *figure 10* – the entire image area is uniformly illuminated so that light is reflected from the whole page. A lens and aperture system scans across a line, and the photodetector receives light from each pixel because the aperture limits the light that can reach the detector. Any light from outside this defined area is blocked, and will not be 'seen' by the detector.

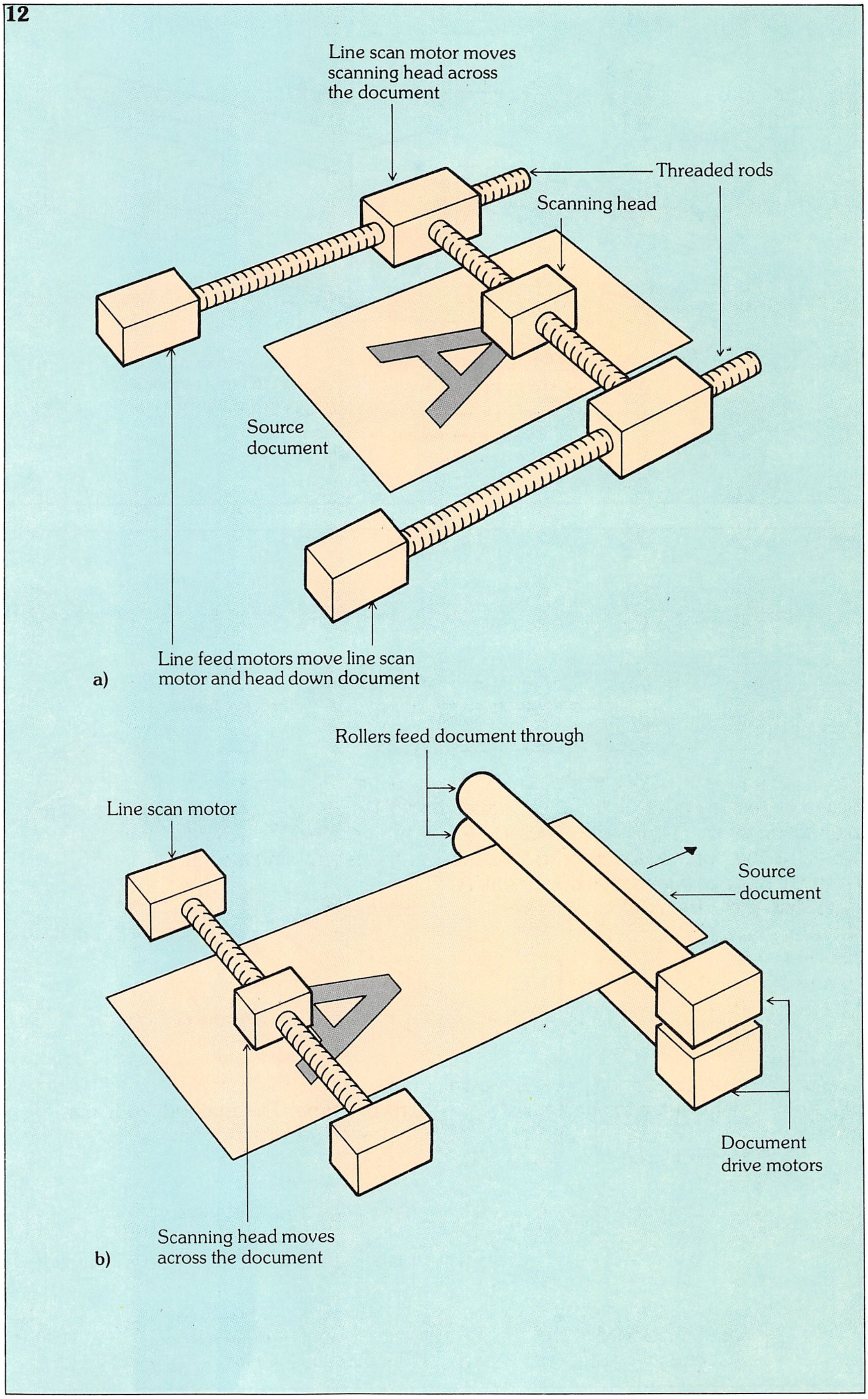
So far we have looked at two methods of detecting light reflected from a document, and two methods of limiting the reflected light in order to resolve the detail of the image. We will now examine the various scanning mechanisms.

Scanners

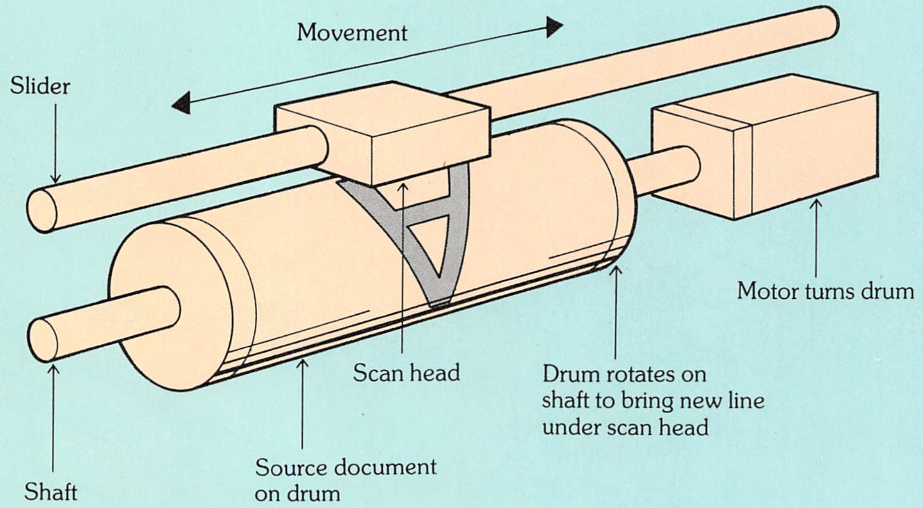
Many techniques have been developed for sweeping or stepping a dot or aperture across the width of the page, and down its length. The simplest technique is to scan a document line by line, in the same way that a television picture is traced out on the face of the CRT – from left to right, one line at a time. This method, known as **raster scanning**, is explained in more detail in *Solid State Electronics 23*.

The basis of raster scanning is, however, that the surface of the source document is broken down by the scanning mechanism into thousands of individual light level readings, each taken over a small dot size. This is illustrated in *figure 11a* for circular dots, and *figure 11b* where square dots are used. If circular dots are used,

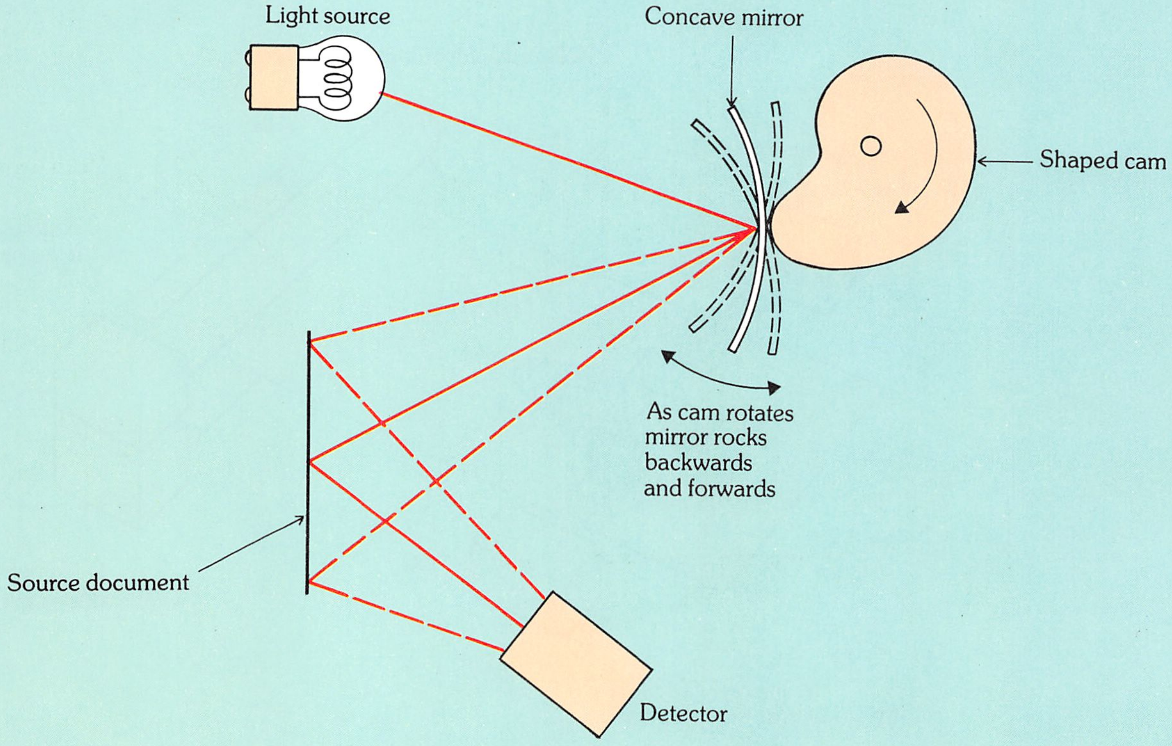
12. **(a)** and **(b):** two flat-bed raster scanning systems.



13

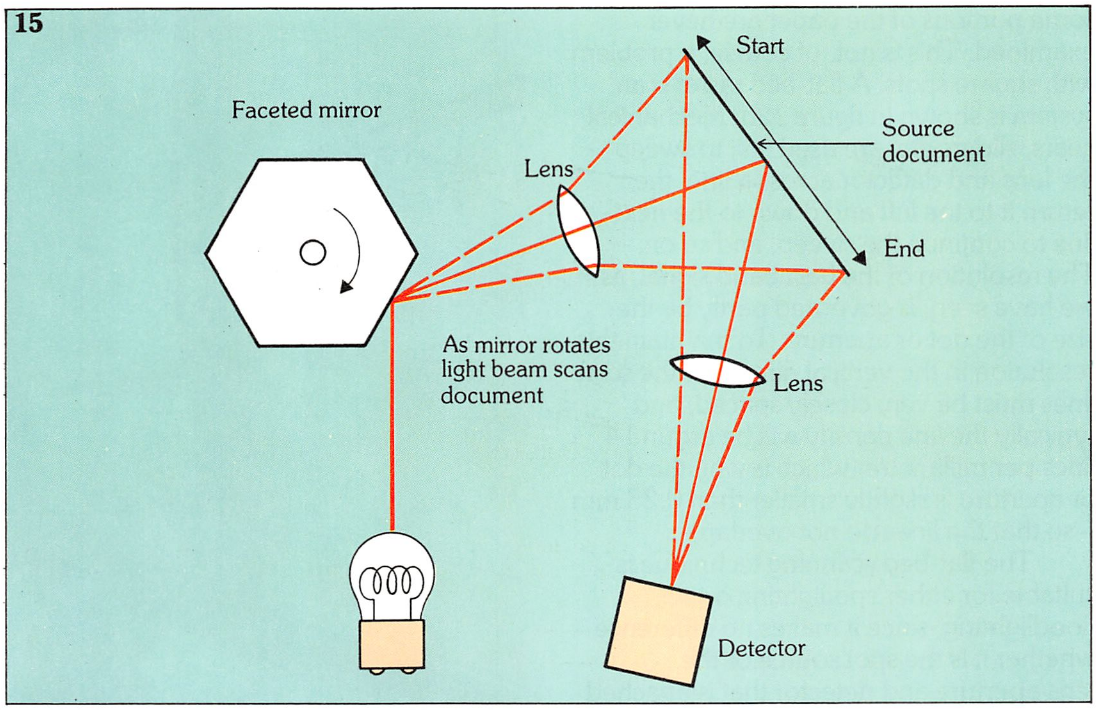

Movement
Slider
Motor turns drum
Scan head
Drum rotates on shaft to bring new line under scan head
Shaft
Source document on drum

14


Light source
Concave mirror
Shaped cam
As cam rotates mirror rocks backwards and forwards
Source document
Detector

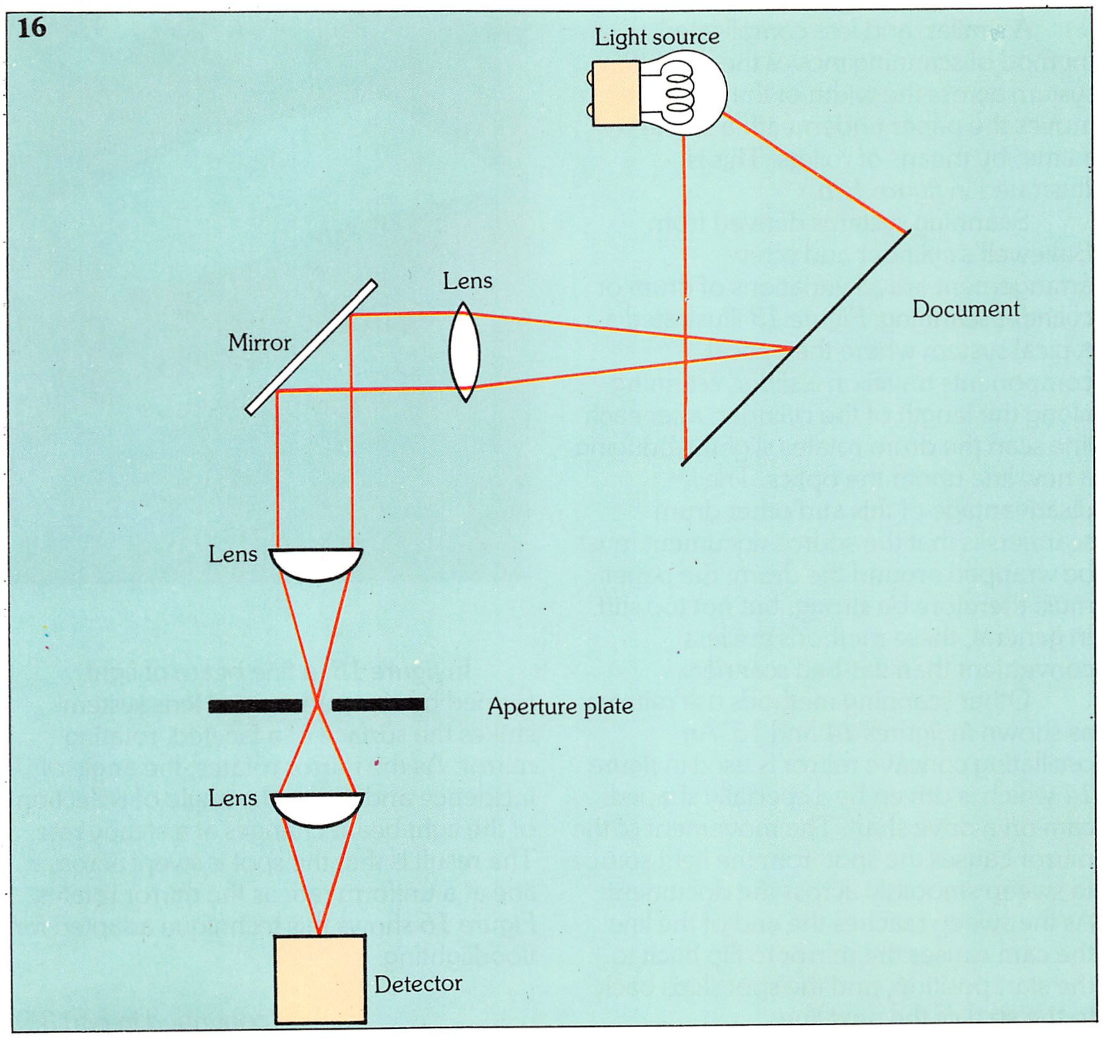
13. Scanning system derived from Bakewell's cylinder and screw arrangement, where the optical components travel on a slider.

14. Scanning system using an oscillating concave mirror.



15. Scanning system using a faceted rotating mirror.

16. The technique shown in *figure 15*, adapted for floodlighting.



some portions of the paper are never examined. This is not, of course, a problem with square spots. A flat-bed raster scan system is shown in *figure 12a*. Mechanical gears, sliders etc. are used first to sweep the lens and detector across a line, then return it to the left and down to the next line to continue the sweep, and so on. The resolution of the baseband signal, as we have seen, is governed partly by the size of the dot or aperture. To maintain this resolution in the vertical direction, the scan lines must be very closely spaced, and typically the line density will be around 4 lines per millimetre (which is why the dot or aperture is slightly smaller than 0.25 mm – so that the lines do not overlap).

The flat-bed scanning technique is suitable for either spotlighting or floodlighting, since it makes no difference whether it is the spot source or the lens/aperture and detector that is attached to the moving carriage.

A similar, and less complicated, method of scanning moves the detection system across the width of the line, but moves the paper underneath a line at a time, by means of rollers. This is illustrated in *figure 12b*.

Scanning systems derived from Bakewell's cylinder and screw arrangement are all variations of drum or cylinder scanning. *Figure 13* illustrated a typical system where the optical components travel on a slider, scanning along the length of the cylinder; after each line scan the drum rotates slightly, bringing a new line under the optics. The disadvantage of this and other drum scanners is that the source document must be wrapped around the drum; the paper must therefore be strong, but not too stiff. In general, these methods are less convenient than flat-bed scanners.

Other scanning methods use mirrors as shown in *figures 14* and *15*. An oscillating concave mirror is used in *figure 14* which is driven by a specially shaped cam on a drive shaft. The movement of the mirror causes the spot from the light source to sweep smoothly across the document. As the sweep reaches the end of the line the cam causes the mirror to flip back to the start position, and the spot skips back to the start of the next line.

Above: British Telecom's Bureaufax being used to send weather maps. This is an international service linking more than 70 different countries.

In *figure 15*, a fine beam of light formed by an aperture and lens system strikes the surface of a faceted, rotating mirror. As the mirror rotates, the angle of incidence and hence the angle of reflection of the light beam changes at a steady rate. The result is that the spot is swept across a line at a uniform rate as the mirror rotates. *Figure 16* shows this technique adapted for floodlighting.

(continued in part 35)